

A Quantitative Method to Study the Relationship between Urban Form and City Liveability Indexes

Alessandro Venerandi

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Engineering
of
University College London.

Department of Civil, Environmental, & Geomatic Engineering
University College London

August 7, 2017

I, Alessandro Venerandi, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Defining the attributes of urban form which relate to city liveability has long been the research topic of a wide range of experts such as architects, urbanists, geographers, and, more recently, computational social scientists. Both qualitative and quantitative approaches have been developed to study this matter; however, the former often lack of generalizability, as results are mainly based on personal views, and replicability, as no systematic methodology has ever been presented. The latter, although being more generalizable, are still geographically constrained to relatively small regions (e.g., a neighbourhood, a city). Moreover, they focus on single aspects of urban form (e.g., density) rather than on multiple ones (e.g., accessibility, density, connectivity), as Urban Morphology predicates. In this thesis, I propose a quantitative approach to study the relationship between multiple aspects of urban form and city liveability indexes that is replicable and applicable to areas of arbitrarily large size. Metrics of urban form are derived from urban theories and extracted from openly accessible datasets such as census data and OpenStreetMap (OSM). These metrics are then used as independent variables in a linear regression model with a liveability index as dependent one. To test the proposed approach, I apply it to different urban regions of the United Kingdom (UK) to understand the relationship between urban form and different aspects of city liveability such as socio-economic deprivation, life expectancy, and childhood obesity. Models show adjusted R^2 values up to 0.76, suggesting good model fit overall. Interpretations of model outcomes and regression coefficients, for the specific geographic context of the UK, suggest that neighbourhoods with worse liveability are characterised by tower block developments, low connectivity, and a predominantly regular street

layout. Conversely, more liveable neighbourhoods tend to be characterised by more connectivity, a denser urban fabric, and an above-average presence of historic buildings.

Impact Statement

Many theories have been formulated throughout human history on what is the urban form that promotes the liveability of cities. Both qualitative and quantitative works have been carried out. However, the former lack of generalizability and replicability, while the latter lack of complexity as they focused on single aspects of urban form. In recent times, the interest in these matters has further grown due the fast urbanisation process that is taking place in cities all over the world. However, as for today, a systematic and quantitative method to analyse the relationship between multiple aspects of urban form and the liveability of cities, for areas of arbitrarily large size, is missing.

To solve this issue, I propose a quantitative methodology, based on regression analysis, to study the relationship between a set of features of urban form and liveability indexes. This method is easily replicable, as it is based on the use of openly accessible datasets. The application of this methodology to a set of UK cities showed that it can be easily implemented and can provide useful insights on to what extent specific features of urban form are related to levels of city liveability.

The proposed methodology can both have theoretical and practical impacts. It can be used by researchers and urban designers to advance knowledge in the design of cities. For example, it can be used to understand the relationship between the built environment and liveability, across different geographic, cultural, and temporal contexts. It can also be applied to understand how the same relationship varies when areas undergo different urban processes, such as gentrification or recession. It can be used by institutions and research bodies to influence building policies. From a practical standpoint, the method can be implemented to build a ‘neighbourhood

profiling' tool, which can be used by different stakeholders. For example, future city dwellers might use it to evaluate different parts of a city and decide where to buy or rent a property; visitors might consult it to decide where to stay; finally, city planners and administrators might use it to analyse and compare what makes a more liveable vs. a less liveable neighbourhood in their cities.

Acknowledgements

This thesis would have not seen the light without the help and support of many. This section is dedicated to them.

I would like to thank my supervisor Licia Capra, for her constant and tireless support and supervising work, for helping me to develop my own research without imposing choices or restrictions, and for allowing me to better understand the efforts needed, but also the high rewards, of academic research.

I would also like to thank my colleague and friend Giovanni Quattrone, a precious guide in the perilous and obscure world of programming, statistics, and databases. I also thank him for being a great friend and for having shared with me lunches, dinners, trips, and research ideas.

Thanks to Daniele Quercia and Diego Saez-Trumper for collaborating on my first paper and for providing indispensable knowledge and valuable tools to advance my research.

Thanks to the Centre for Urban Sustainability and Resilience (USAR), to Helena Titheridge, Jane Doogan, and Pavlina Pyari, for being always responsive, for providing guidance, and for helping with administrative tasks, in the fastest and smoothest way possible. A special thank goes to Simon Justin of the IT team, for lending me his laptop, after mine broke without remedies few months before the end of the doctorate.

Thanks to the Engineering and Physical Science Research Council (EPSRC) for having funded my research.

I thank my mum and dad, for their constant affection and support in these long years that saw me far from my birthplace. They were distant but very close to my

heart. I hope they felt the same.

Finally, I would like to thank my girlfriend, Iana, whose constant affection, support, and presence were indispensable to get along with life and work. Also, thanks for still having the patience to correct my English.

*a mia madre,
che mi ha donato amore incondizionato e che sempre rimarrà nel mio cuore*

Contents

1	Overview	21
1.1	Hypotheses	23
1.2	Thesis' Contributions	23
1.3	Thesis Outline	24
1.4	Publications	25
2	Literature review	27
2.1	Liveability and Ways to Measure It	27
2.2	Urban Theories	29
2.2.1	Ancient Times	29
2.2.2	Modern Era	30
2.2.3	Quantitative Urban Studies	33
2.3	Beyond Liveability	38
2.3.1	Transport	38
2.3.2	Social Science	39
2.3.3	Geography	40
2.3.4	History	40
3	A Quantitative Method to Study Urban Form and City Liveability	43
3.1	Extraction of Metrics	44
3.2	City Liveability Indexes	47
3.3	Analytical Approach	47
3.3.1	Spatial Unit of Analysis	48

3.3.2	Correlation, Regression Analysis, and Interpretations	49
-------	---	----

4	On the Relationship Between Amenities and Socio-economic Deprivation	53
4.1	Introduction	53
4.2	Datasets	54
4.2.1	Index of Multiple Deprivation	56
4.2.2	Foursquare	59
4.2.3	OpenStreetMap	60
4.3	Approach	62
4.3.1	Metrics	62
4.3.2	Spatial Unit of Analysis	63
4.3.3	Correlation Analysis	64
4.3.4	Regression Analysis	65
4.4	Results	68
4.4.1	Amenities and Deprivation	68
4.4.2	Modelling Amenities and Deprivation	71
4.4.3	Interpretations	73
4.5	Discussion	77
4.5.1	Limitations	77
5	On the Configuration of the Urban Environment and Socio-economic Deprivation	81
5.1	Introduction	81
5.2	Datasets	82
5.3	Approach	84
5.3.1	Metrics	87
5.3.2	Spatial Unit of Analysis	92
5.3.3	Regression Analysis	92
5.4	Results	94
5.4.1	Preliminary results	97

5.4.2	Modelling Urban Form and Deprivation Across the Six Cities	98
5.4.3	Interpretations	100
5.5	Discussion	103
5.5.1	Limitations	104
6	On the Relationship between Amenities, Configuration of the Urban Environment and Three Liveability Indexes	107
6.1	Introduction	108
6.2	Datasets	109
6.3	Approach	111
6.3.1	Metrics	111
6.3.2	Spatial Unit of Analysis	111
6.3.3	Correlation Analysis	112
6.3.4	Regression Analysis	112
6.4	Results	114
6.4.1	Amenities, Configuration of Urban Form and IMD	114
6.4.2	Interpretations	115
6.4.3	Amenities, Configuration of Urban Form, Life Expectancy and Childhood Obesity	117
6.4.4	Interpretations	121
6.5	Discussion	122
6.5.1	Limitations	123
7	General Conclusions	125
7.1	Summary of Contributions	125
7.1.1	Amenities and Socio-economic Deprivation	126
7.1.2	Configuration of Urban Form and Socio-economic Deprivation	126
7.1.3	Amenities, Configuration of Urban Form and Three Liveability Indexes	127
7.2	Discussion	128

7.2.1	Engineering	128
7.2.2	From Model to Prediction	130
7.3	Implications	131
7.4	Limitations	132
7.5	Future Work	134
Appendices		137
A	A sample of the Index of Multiple Deprivation (IMD) dataset	137
B	Maps of the configuration of the urban environment and deprivation for the six urban areas under study	139
Bibliography		146

List of Figures

3.1	Flow chart of the methodology.	45
4.1	Location of the urban areas under study in the UK.	55
4.2	Maps of deprivation of the urban areas under study.	57
4.3	Foursquare amenities in central areas of the urban areas under study.	58
4.4	OSM amenities in central areas of the urban areas under study.	61
5.1	Location of the six urban areas under study.	83
5.2	Building footprints and road segments in central areas of Newcastle, Leeds, and Greater Manchester.	85
5.3	Building footprints and road segments in central areas of Liverpool, West Midlands, and Greater London.	86
5.4	Density distributions of the metrics of urban form and IMD.	95
6.1	Life expectancy scores for Greater London.	110
6.2	Childhood obesity scores for Greater London.	111
B.1	Maps of the metrics of the configuration of the urban environment and deprivation for Newcastle.	140
B.2	Maps of the metrics of the configuration of the urban environment and deprivation for Leeds.	141
B.3	Maps of the metrics of the configuration of the urban environment and deprivation for Greater Manchester.	142
B.4	Maps of the metrics of the configuration of the urban environment and deprivation for Liverpool.	143

B.5	Maps of the metrics of the configuration of the urban environment and deprivation for West Midlands.	144
B.6	Maps of the metrics of the configuration of the urban environment and deprivation for Greater London.	145

List of Tables

2.1	Main urban theories, principles, and relative authors.	36
3.1	Examples of transformations of qualitative theories into quantitative metrics.	46
4.1	Population, area, and density of the urban areas under study.	54
4.2	General information on the datasets used in the analysis.	56
4.3	Number of Foursquare amenities, check-ins, and categories across the urban areas under study.	59
4.4	Number of OSM amenities and categories across the urban areas under study.	62
4.5	Number of Foursquare categories of amenities correlated with IMD.	69
4.6	Number of OSM categories of amenities correlated with IMD.	69
4.7	Foursquare categories of amenities most correlated with IMD.	70
4.8	OSM categories of amenities most correlated with IMD.	70
4.9	Outcomes of the False Discovery Rate test.	71
4.10	LR and SAR models for Greater London.	72
4.11	LR and SAR models for Greater Manchester.	73
4.12	LR and SAR models for West Midlands.	74
5.1	Population, area, and density of the urban areas under study.	82
5.2	Number of road segments and building footprints for the six urban areas under study.	84
5.3	Metrics of the configuration of the urban environment.	91

5.4	Mean and standard deviation values for the metrics of the configuration of the urban environment and IMD.	93
5.5	LR and SAR models for the six urban areas under study.	101
6.1	Metrics of amenities and configuration of the urban environment. . .	112
6.2	LR and SAR models for IMD.	115
6.3	LR and SAR models for life expectancy.	117
6.4	LR model for the residuals of the regression between income deprivation and life expectancy.	118
6.5	LR and SAR models for childhood obesity.	119
6.6	LR model for the residuals of the regression between income deprivation and childhood obesity.	119
7.1	Performance of the classifier for Greater London.	130
A.1	Sample of the 2011 IMD dataset.	137

Chapter 1

Overview

Urban form is the set of physical features, for example roads and buildings, that characterise urban settlements [1, 2]. The relationship between urban form and the liveability of cities has for long been one of the most relevant research topics in architecture and urban design. It was hypothesised that certain configurations of the built environment were more related to better outcomes in terms of liveability than others, and that this would have ultimately influenced the social, cultural, and economic development of the entire society [3, 4, 5, 6].

This topic has recently gained even more recognition due to current urbanisation, environmental changes, and socio-economic trends. On the one hand, the fast and unprecedented urban growth, that is likely to bring 68% of the world total population to cities by 2050 [7], has made urgent acquiring more understanding on the relationship between the way we build cities and the well-being of residents. On the other hand, climate change and serious environmental issues require considering this relationship in light of sustainability and resilience so that urban form will not only bring prosperity to human beings but also to the whole ecosystem. Finally, social equity has to be respected if we want prosperity to be shared rather than benefit a few. In this regard, recent research reported that inequality is dangerously on the rise worldwide, with certain regions benefiting more than others of public fundings and economic growth [8]. All of these matters have to be considered when analysing urban form in relation to liveability.

Nonetheless, as for today, a rigorous and replicable method to investigate such

relationship is missing. Past qualitative urban theories provided useful insights; however, the methodology adopted was based on personal views or qualitative observations. Such approach was very costly and thus difficult to replicate across different contexts. Furthermore, these studies were usually bounded to small geographic areas and rarely replicated over time. More recent works do present quantitative approaches; however, they tend to study single aspects of urban form in relation to aspects of well-being. For example, Hillier studied the relationship between cul-de-sac in street networks and presence of crime [9]. However, the complexity of spatial structures requires a more holistic approach and the analysis of multiple aspects at the same time. The discipline of Urban Morphology supports this and considers cities in light of their multiple basic components (e.g., streets, buildings, blocks) [1, 2] rather than single city-wide phenomena, such as the distribution of functions at the regional scale [10], networks [11], size/performance of cities [12], or vast urbanisation processes [13].

The lack of a quantitative method that supports the systematic analysis of urban form in relation to city liveability could have been due to the difficulty of acquiring data. In recent years, though, the production of data of any kind has significantly increased. This phenomenon is referred to as “open data revolution” [14, 15], an expression which emphasises both the large quantity of information involved and its novelty aspect. Never before, in fact, it has been possible to produce and access such a vast amount of data. Among this, one can find information pertaining to the built environment such as roads, amenities, and buildings as well as data regarding levels of liveability. What is missing then is a methodology to study these two in relation to one another.

In this thesis, I propose a quantitative method, based on Urban Morphology theory, that extracts multiple metrics of urban form from a variety of openly accessible datasets, to study the relationship between the built environment and city liveability. This method can be applied to urban regions of arbitrarily large size, to test a wide variety of research questions related to socio-economic aspects, well-being, and sustainability. Moreover, it can be largely automated so that it can be

easily replicated over time, to observe possible changes in these relationships, for example, as cities undergo processes of further urbanisation or gentrification.

1.1 Hypotheses

The formulation of the methodology is based on two main hypotheses. These are derived from previous studies and theories on the association between features of urban form and aspects of city liveability. These are:

Hypothesis #1: whether or not specific amenities are associated with levels of city liveability. Such amenities can be restaurants but also specific public facilities (e.g., school, museum) or religious buildings (e.g., mosque). This hypothesis is inspired by the findings of studies in the field of Preventive Medicine. For example, Giles-Corti and Donovan found that higher concentrations of golf courses in Australia were related to higher socio-economic statuses [16]. Cummins *et al.* reported that higher densities of fast food restaurants in England and Wales were associated with more deprived neighbourhoods [17].

Hypothesis #2: whether or not aspects of the configuration of the built environment are related to levels of city liveability. These aspects relate to the shape and properties of the urban fabric (e.g., built density, connectivity). This second hypothesis is inspired by the outcomes of quantitative works in the field of Urban Design. For example, Vaughan found that more street network accessibility was associated with better-off residents, in an East London area [18]. Hillier and Sahbaz reported that density was overall beneficial against crime, in a London borough [19].

1.2 Thesis' Contributions

Based on the previous hypotheses, this thesis develops the following contributions:

- a quantitative method to study the relationship between urban form, as composed by amenities and configurational aspects of the built environment, and indexes of city liveability through the use of ready available, openly accessible datasets and statistical analysis. Moreover, this method affords the study of urban regions of arbitrarily large size, for different time frames;

- five applications of this method to analyse the relationship between urban form and different aspects of city liveability, in six urbanised areas of the UK. More precisely, the aspects analysed are: socio-economic deprivation, life expectancy, and childhood obesity;
- interpretations for the quantitative outcomes, in light of previous urban theories and related works.

1.3 Thesis Outline

The reminder of this thesis is structured as follows:

Chapter 2 presents an overview of the most relevant qualitative and quantitative contributions concerning urban form and liveability of cities. It then critically challenges their outcomes and methodologies.

Chapter 3 illustrates the core subject of this thesis, that is a quantitative methodology to study the relationship between urban form and aspects of liveability of cities.

Chapter 4 focuses on the testing of the first hypothesis of this thesis (i.e., amenities are related to aspects of city liveability) by applying the proposed methodology to levels of socio-economic deprivation of three UK major cities.

Chapter 5 focuses on the testing of the second hypothesis of this thesis (i.e., configurational aspects of the built environment are related to aspects of city liveability) by applying the proposed methodology to levels of socio-economic deprivation of six UK major cities.

Chapter 6 focuses on the testing of both hypotheses together by applying the proposed methodology to both socio-economic and health scores of London neighbourhoods.

Chapter 7 presents conclusive remarks concerning achievements and limitations of the proposed method and its applications, practical uses of the method, and future work.

1.4 Publications

The publications listed below have been produced while developing this thesis:

Venerandi, A., Quattrone, G., Capra, L., Quercia, D., and Saez-Trumper, D. “Measuring urban deprivation from user generated content.” In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing. ACM, 2015.

Venerandi, A., Quattrone, G., and Capra, L. “Guns of Brixton: which London neighborhoods host gang activity?.” In Proceedings of the Second International Conference on IoT in Urban Space. ACM, 2016. *Best paper award*.

Venerandi, A., Quattrone, G., and Capra, L. “City form and well-being: what makes London neighborhoods good places to live?.” In Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2016.

Venerandi, A., Quattrone, G., and Capra, L. “A scalable method to quantify the relationship between urban form and socio-economic indexes.” Submitted to EPJ Data Science.

Chapter 2

Literature review

The relationship between urban form and the liveability of an urban environment has been investigated in a variety of different ways and from multiple perspectives throughout history. Works related to this topic can be broadly subdivided in two groups, qualitative and quantitative works, depending on the methodological approach used. I start by introducing the concept of liveability and some examples of how it is computed. I then present the works that adopted a qualitative approach, their achievements and limitations. Finally, I illustrate studies that adopted a quantitative approach, their achievements and limitations.

2.1 Liveability and Ways to Measure It

The concept of liveability is strongly associated with quality of life and thus refers to the overall well-being of individuals and communities. This does not only include economic aspects, such as income and employment, but also health conditions, the possibility of having leisure time, and social belonging [20]. Other common concepts associated with liveability and quality of life are personal freedom, human rights, and happiness. However, since these last aspects are more difficult to quantify as there are no objective metrics to do so, governments and institutions generally paid less attention to them and focused more on economic and health aspects, instead.

Two well known measures of liveability are the *Where-to-Be-Born Index* issued by the Economist Intelligence Unit [21] and the *Quality of Living* published

by Mercer [22], a consulting firm. Both measures are computed at country level by considering several different aspects of people's lives, such as life expectancy, job security, social rights, and gender equality. The *Human Development Index (HDI)*, is another index of liveability, at country level, developed by the United Nations (UN). The HDI is calculated by combining life expectancy, standard of living, and education to quantify the chances that a person has in a specific society [23]. The rankings obtained from these metrics generally provide similar results, with the most commonly ranked countries in the top 50 being Austria, Switzerland, New Zealand, Germany, Canada, Australia, Sweden, and the US. The *Happy Planet Index* is an alternative index, instead. This is computed not only by including aspects of well-being, but also the ecological footprints of countries [24]. Western countries thus do not occupy the first positions as they tend to pollute more. The 2012 index saw Costa Rica, Vietnam and Colombia at the top of the ranking. The Bhutan *Gross National Happiness (GNH)* measure is another example of alternative index as it focuses on the subjective happiness of individuals rather than economic aspects [25]. The indexes presented in this paragraph provide insightful information on different levels of liveability of countries. However, this information is provided at a coarse level of spatial granularity.

There exist others, though, that are computed for smaller spatial units and that thus provide information on liveability for regions within countries. For example, the UN *Multidimensional Poverty Index (MPI)* is computed for geographic clusters of 5,000 people and covers over 100 developing countries [26]. The MPI extends the more income-based HDI by including data on education, health and living standards. The UK *Index of Multiple Deprivation (IMD)* is another example of liveability measure computed at a fine level of spatial granularity [27]. IMD is calculated for small census areas of around 1,500 residents by considering both income-related information and aspects of health and education. However, the former have more statistical weight. Since these indexes are computed for smaller spatial units, they offer the opportunity to investigate urban form at this scale. I present the methodology to do so in Chapter 3, while I illustrate theories and quantitative studies on

urban form and liveability in the next section.

2.2 Urban Theories

2.2.1 Ancient Times

The search for an urban form that promotes liveability is a long standing goal which is perhaps as old as civilisation itself. Although history attributes to an Ancient Greek – Hippodamus of Miletus (V century BC) – the first ever generated theory of urban form, that of the grid layout, there is evidence supporting a more ancient dating for this specific way of planning. Older archaeological remains, in the form of house platforms and foundations, from the Greek colonies of Agrigento, Metaponto, and Selinunte, all of which are located in Italy, seem to confirm that the grid street layout was already in use [28]. However, no written codification of such way of planning has been found so far. For what concerns the Hippodamian city, it was based on a grid with three to four main longitudinal roads intersected by minor streets at right angles, whose aim was only to delimit blocks that, for this reason, were long and thin. Their short side was often set at 36.6 meters. The main roads did not meet in a central space like in Roman planning. Nevertheless, squares were present in the plan [29].

After Hippodamus, others architects provided their visions for an ideal city. Roman architect and engineer Vitruvius (I century BC), for example, theorised a city with a central space at its core, surrounded by several public buildings such as a basilica, a treasury, and the senate floor. This city was shaped as an octagon, according to the eight main wind directions. This specific design, in Vitruvius' view, would have avoided strong winds to hit the streets directly [30].

More recently, during the Renaissance period, the Italian architect Vincenzo Scamozzi (XVI - XVII century AC) proposed a circular shaped city surrounded by walls with a central square at its core. This, like in Vitruvius' plan, was surrounded by many important buildings such as the monarch residency, a cathedral, a finance office, and a law court. The street layout was based on the grid. However, variations in the size of blocks were present. These were due to the circular shape of the city

and the presence of several other squares [31].

2.2.2 Modern Era

Fast forward to more recent times, the car as a mode of transport for the masses has been the most important paradigm upon which the debate on the optimal form of cities has been centred. The dilemma was between designing cities for humans, where cars were allowed while not being the main actors, or shaping the urban environment just for cars, for example, with super highways cutting in half city centres or huge parking lots subtracting space to buildings and amenities. In the second half of the twentieth century, these opposite views gave birth to two different schools of thoughts. On one side, there were those supporting the compact, highly walk-able city form; on the other, those favouring a more car-oriented urban environment. Journalist and activist Jane Jacobs (1916 - 2006) belonged to the first group and was a strenuous supporter of the traditional compact city form. In her view, a “good” city was characterised by medium to high population and built density, blocks facing the streets (perimeter blocks), walk-ability, mix of different uses and functions, and mix of buildings of different ages. Moreover, she identified three qualities for a safe urban environment. There had to be clear demarcations between public and private domains, buildings should face the streets so that their inhabitants can see what is happening on them (i.e., “eyes on the street” effect), footpaths should constantly be frequented by passers-by to increase the level of informal visual control. For what concerns the street layout, Jacobs favoured the grid plan if this was to be characterised by variations such as squares and diagonal roads [3]. Similarly to Jacobs, American urbanist William Hollingsworth Whyte (1917 - 1999) supported the traditional city form by praising walking, the human scale size of streets and buildings, and those subtle but indispensable urban features such as curbs, footpaths, shop windows, doorways, porticoes, and steps [32]. Fast forward to present days, Danish architect Jan Gehl (born 1936) is at the forefront of promoting the dense city, not only from a theoretical standpoint but also from a very practical one, through design and construction. Similarly to the previously mentioned authors, he favours human scale, pedestrian, mixed use streets and attributes

great importance to active *city edges* (block frontages). These should be characterised by shop windows, doorways and forecourts so to promote social interactions [4]. Among his many built projects, here I mention one which strongly embodies the concepts of his theory. This is the redesign of New Road in Brighton (UK) which saw the transformation of this central street from a mere transport link to a public space with a wide newly paved shared surface where people can just stroll, sit on a beautifully crafted long wooden bench, visit the shops, or go to the many theatres located on the street.

As I mentioned before, there were two different schools of thoughts regarding city form: one supported the traditional compact city form while the other a more dispersed and car-oriented urban form. Architect Charles-Edouard Jeanneret-Gris (1887 - 1965), better known with the pseudonym of Le Corbusier, belonged to the latter group. His view on planning was dominated by the car and technology in general. He negatively criticised the dense historic city claiming it was disordered, chaotic, and unhealthy and proposed a new way of planning based on standards and efficiency. Given the advent of the car, he alleged that walking was no longer necessary and that cities should therefore not be built dense any more, but dispersed as the eventual distance between places could be covered by car. Following this idea, he proposed a city characterised by super blocks delimited by multi lane highways with residential towers detached from streets and laid out in undefined green spaces (i.e., the so called “towers in the park”). The other recurrent characteristic of his plans was the separation of functions (e.g., residence, work, leisure) in dedicated areas (this approach is also known as zoning, that is the separation of functions in specific zones) [5]. Ludwig Hilberseimer (1885 - 1967) and Walter Gropius (1883 - 1969) were also supportive of this approach. The former emphasised the new technical achievements of the car and the recently discovered reinforced concrete, by proposing city plans characterised by vast highways and repetitive, tall blocks of flats aligned to an orthogonal street grid [6]. Gropius, beside being in favour of the above mentioned features, supported the assumption that the social patterns in old city centres were disintegrating and thus, following the credo of having sepa-

rated functions, he suggested the creation of community centres to contravene to this issue [33].

More recently, researchers studied urban form in relation to socio-economic changes that cities undergo, for example in the occurrence of gentrification. Historically, this phenomenon has been defined as the process by which more advantaged social classes (*gentry*) substituted less advantaged ones, in specific city neighbourhoods [34]. In fact, gentrification has shown to be a much more complex process which involves, not only the socio-economic aspect of city dwellers, but also land value, the upgrade of existing housing stock, with its relative amenities and services, and the profile of residents as well as visitors [35]. Assuming there was no unique template for gentrified urban areas, scholars, nevertheless, reported recurrent patterns in urban form where gentrification took place. Pacione, for example, identified several physical characteristics of neighbourhoods associated with this phenomenon: the clustered presence of substandard but structurally sound housing, rare amenities such as views or good transport connections, and the presence of attractive, local commercial activities [35]. Butler *et al.* described gentrified neighbourhoods as dense and vibrant, well-connected to the centre while not being the centre, with a good variety of services and amenities, characterised by terraced houses, or by cottages and mews, or by Victorian houses [36]. Others linked gentrification to large scale developments built on peripheral or central vacant lands by the hand of large corporates or investment firms [37, 38]. Linked to this, Shaw argued that gentrification could embrace different physical connotations, that of the compact urban form, but also that of the high-end residential towers in gated communities [39].

All the works presented in this section were – and still are – extremely influential in research, policy making, design and construction. This relevance is conceivably due to their ability to interpret, with an informed eye, the urban realm, society, technology, and ongoing changes. However, these theories often clashed and supported contrasting point of views, if not opposite. Furthermore, they were usually formulated through the use of qualitative methodologies. For example, they were

based on personal views or direct observation of specific parts of cities. Moreover, they were not repeated over time. The outcomes of these works are thus hardly generalizable and cannot be extended to different geographic contexts and time frames.

More recently, with the advent of Geographic Information Systems (GIS) and new conspicuous data repositories, researchers started to scale up studies on urban form and liveability through the use of quantitative methods. I present next these works, their achievements, and limitations.

2.2.3 Quantitative Urban Studies

To overcome the lack of generalizability of the above mentioned qualitative studies, researchers investigated the adoption of quantitative methods to complement qualitative ones and used the latter as theoretical bases for the formulation of metrics. For example, one of the urban features considered by Jacobs was urban density [3]. Quantitative researchers ‘translated’ this aspect in population density, that is the ratio between the number of residents living in a neighbourhood and its total area. Other researchers used more complex metrics. For example, Vaughan used Space Syntax [40], a renowned technique for the analysis of street networks, to study the relationship between the spatial distribution of social classes and *integration* in East London [18]. This latter metric measures to what extent each road segment in a street network is more accessible (integrated) or less accessible (segregated). Results suggest that more integrated places (e.g., high streets, affluent places) were related to the presence of more well-off residents, while more segregated ones (e.g., back streets, interstitial spaces) were associated with less advantaged classes. Hillier also focused on a specific configurational aspect of the street network, that of the cul-de-sac (or dead-end road), and studied this in relation to crime occurrences in a UK new town [9]. Previous qualitative theories were, in fact, discordant on this specific topic. On one side, Jacobs claimed that a well-connected street network that allowed the passage of many passers-by, offering “eyes on the street”, would have prevented crimes [3]. On the other, Oscar Newman was in favour of cul-de-sac as, in his view, less connectivity and thus less passage of people meant fewer chances of criminal acts [41]. Outcomes from Hillier’s research suggest that cul-de-sac did not

necessarily attract more crimes if they were integrated in a street network with significant through movement and many houses facing the streets. Other scholars focused on how different values of built density were related to crime levels but found discordant results. Some did not find any relationship between the two [42, 43, 44], while a more recent study reported that density was overall beneficial against crime and that the flat was the safest house type in a London borough [19]. Density has also been studied in relation to social aspects of city liveability. However, as for the above mentioned studies, outcomes were contrasting. Some researchers found that high density facilitated social interactions at the street level [45, 46], that low density diminished the chances of spontaneous meetings and induced more car dependent behaviours [47]. Conversely, other scholars reported that high density made people less sociable and increased their levels of stress [48, 49, 50]. The effects of density have been analysed also in relation to place attachment and aesthetics [51, 52, 53]. However, results were discordant. Other works focused on housing typologies and alternative social aspects of city liveability. Researchers in this field found that the flats were associated with lower place satisfaction, more neighbourhood issues, but also with better access to amenities and services [54, 55].

Other features of urban form studied in relation to aspects of liveability were amenities such as restaurants and shops. Some researchers reported that health-promoting facilities (i.e., golf courses in Australia [16], fitness centres and dance schools in the USA [56]) were more related to well-off neighbourhoods. Conversely, other scholars found that potential health-harmful amenities (i.e., fast food restaurants in England and Wales [17]) were more concentrated in less advantaged areas.

The above mentioned studies were mainly carried out by extracting metrics of urban form from datasets specifically collected or from proprietary datasets. These metrics quantified various aspects of urban form (e.g., accessibility, built density, house type) and were studied separately in relation to different aspects of city liveability, through correlation or regression analysis. However, although quantitative methods were implemented, generalizability is still an issue of concern as many re-

sults are discordant. This might be due to the fact that different geographic contexts and time periods were considered. Another common limitation is that features of urban form have been analysed separately. However, urban form is a complex and composite entity, constituted by many interrelated features, which thus requires a multivariate approach and the use of multiple morphological variables rather than single ones.

In the last decade, with the advent of the web 2.0 and new techniques of data collection (e.g., crowd-sourcing), computational social scientists have tried to overcome the above mentioned limitations by (i) developing methods applicable to regions of arbitrary large size thus improving the generalizability of results and by (ii) analysing pictures of city settings which, by default, encompass multiple features of the urban environment. Quercia *et al.* [57], for example, used a crowd-sourcing technique to ask around 3,300 people whether pictures of different urban environments in London transmitted beauty, quietness, and happiness. They then used this information to identify what visual characteristics (e.g., presence of specific colours, textures) were associated with the above mentioned three qualities. They found that the colour green – reflecting the amount of greenery – was the factor most positively associated with all qualities, while visual features interpreted as wide roads, faceless buildings, and council houses were inversely related to those qualities. Salesses *et al.* used a similar methodology (i.e., crowd-sourcing) to ask around 7,900 people their perceptions on the safety, social status, and uniqueness of different urban environments across four cities, in the US and Austria [58]. Unlike Quercia *et al.* who focused on visual features, Salesses *et al.* analysed user ratings in relation to socio-demographic factors. The authors reported that spatial variations for perceptions of safety and class better correlated with violent crimes than their absolute values. Furthermore, they found that, safety perception being equal, these crimes occurred in areas that looked more upper class. In a subsequent study, Naik *et al.* used these findings to develop an algorithm (i.e., Streetscore) that, given in input an image of a street scape, estimated its safety [59].

The works presented in this last paragraph focused on point-level data, that is

points defined by a set of geographic coordinates (i.e., latitude and longitude). The focus of my thesis is on areal data, instead, as elements of urban form (e.g., connectivity of neighbourhoods) are hardly reducible to points. Nonetheless, these works inspired me to study urban form in a more comprehensive manner and thus by considering multiple features rather than only one at a time. This would, for example, involve the study not only of spatial accessibility in relation to socio-economic levels of residents, as in the work by Vaughan *et al.* [18], but spatial accessibility, connectivity, and density in relation to socio-economic levels. I present a summary of the works presented above and their guiding principles at the end of this chapter, in Table 2.1.

Table 2.1: Main urban theories, principles, and relative authors.

<i>Theory</i>	<i>Author</i>	<i>Principle</i>	<i>Method</i>
Grid layout	Hippodamus of Miletus [29], Vitruvius [30], Scamozzi [31]	Orthogonal system of streets with few exceptions (e.g., squares, public buildings)	Qualitative
Compact city form	Jacobs [3], Whyte [32], Gehl [4]	Medium to high population and built densities, perimeter blocks, walk-ability, mixed-use	Qualitative
“Towers in the park”	Le Corbusier [5], Gropius [33], Hilberseimer [6]	Isolated residential towers surrounded by open space and major arteries, separation of functions (i.e., zoning)	Qualitative
Gentrification	Pacione [35], Butler <i>et al.</i> [36]	Set of socio-economic as well as physical changes that affects parts of cities with the following characteristics: historic housing, variety of amenities, rare amenities, good connectivity to the centre	Qualitative

Continued on next page

Table 2.1 – Continued from previous page

<i>Theory</i>	<i>Author</i>	<i>Principle</i>	<i>Method</i>
Space Syntax	Vaughan <i>et al.</i> [18], Hillier [9]	More accessible places are associated with better socio-economic conditions	Quantitative
Built density is beneficial against crime	Hillier & Sahbaz [19]	Denser urban environments are beneficial against crime. The flat is the safest house type	Quantitative
	Haughey [42], Harries [43], Li & Rainwater [44]	There is no relationship between density and crime	Quantitative
Built density is beneficial to social interactions	Duany <i>et al.</i> [45], Talen [46], Burchell <i>et al.</i> [47]	Density facilitates social interactions at the street level. Lower densities reduce them and lead to car-dependent behaviours	Quantitative
	Wirth [48], Simmel [49], Freeman [50]	Density makes people less sociable and increases levels of stress	Quantitative
Specific house types are better than others	Bramley <i>et al.</i> [54], Bramley & Power [55]	Flats are associated with lower place satisfaction, more social issues, but better access to amenities	Quantitative
Specific amenities are associated with socio-economic levels	Giles-Corti & Donovan [16], Powell <i>et al.</i> [56]	Health-promoting facilities (i.e., golf courses, fitness centres, dance schools) tend to be more present in better-off neighbourhoods	Quantitative
	Cummins <i>et al.</i> [17]	Health-harmful amenities (i.e., fast food restaurants) tend to be more present in more deprived neighbourhoods	Quantitative

Continued on next page

Table 2.1 – Continued from previous page

<i>Theory</i>	<i>Author</i>	<i>Principle</i>	<i>Method</i>
Visual features of streets are associated with positive psychological statuses	Quercia <i>et al.</i> [57]	Colour green is associated with beauty, quietness, and happiness of places. Wide roads, faceless buildings, council houses are inversely related to them	Quantitative
Visual perceptions are associated with socio-demographic factors	Salesses <i>et al.</i> [58]	Spatial variations in perception of safety and social class are correlated with violent crimes. Violent crimes occur in areas that look more upper class	Quantitative

2.3 Beyond Liveability

Urban form is not the only factor influencing city liveability and its multiple aspects (e.g., well-being, socio-economic levels). Many other phenomena and dynamics play important roles as well. These factors are the subjects of a wide range of disciplines, for example, Demographics, Sociology, and Geography. While it would be impossible to present in this thesis a complete overview of the many phenomena linked to city liveability and relative disciplines, I provide a selection of the ones that more markedly are related to urban form.

2.3.1 Transport

Researchers in this field analysed, for example, the relationship between the flow of commuters using the London underground train service and the socio-economic levels of neighbourhoods. To this end, Smith *et al.*, extracted metrics of mobility (i.e., flow between areas and choice of transport mode) from Oyster Card data (i.e., the electronic ticketing system in use in London) to build a classifier of socio-economic deprivation [60]. The model showed an overall good performance as it was able to explain deprivation in a binary fashion (i.e., below and above the median value) with

a precision of 0.805. However, this result applied only to a relatively small subset of London neighbourhoods, the ones with underground stations (i.e., 10%). Nevertheless, the outcomes of this study confirm the existence of a relationship between an aspect related to transport and socio-economic levels. Simiraly, Lathia *et al.* analysed the relationship between mobility flows of commuters in London, as recorded via Oyster cards, and the socio-economic deprivation of neighbourhoods with underground stations [61]. Outcomes suggest that such correlation exists. Moreover, the authors found that less deprived areas tended to attract people coming from neighbourhoods with various deprivation scores, while more deprived ones tended to attract people coming from other deprived areas, indicating social segregation.

2.3.2 Social Science

Records of calls and text messages, also called “Call Detail Records” (CDRs), exchanged through telecommunication devices, have been analysed in relation to the socio-economic levels of different populations around the world. Researchers used the information stored in these records (e.g., time, duration, caller ID, callee ID and the location of the antenna through which the communication took place) to derive metrics and correlate them with socio-economic indexes. One of the first studies in this field has been carried out by Eagle *et al.* [62]. This group of researchers studied the relationship between CDRs of land lines and mobile phones in England and an official index of socio-economic deprivation and reported a strong correlation between a measure of diversity of contacts and higher socio-economic levels. In a more recent study, Mao *et al.* analysed whether features of calls in Cote d’Ivoire were related to the economic indicators of ten regions with a strong economic activity [63]. They reported that the ratio between the outgoing calls in each area and the total of outgoing plus incoming calls had a strong correlation with income per year. Smith *et al.* analysed the same dataset and found that measures of network diversity and introversion (i.e., ratio between within-area calls and inter-area calls) were strongly associated with socio-economic deprivation [64].

2.3.3 Geography

Researchers in the field of Geography and Remote Sensing analysed whether patterns of the Night Time Light (NTL) (i.e., the amount of Earth's surface lit at night time), extracted from satellite images, were associated with the economic development of several countries. Elvidge *et al.* reported a relationship between NTL and country level Gross Domestic Product [65, 66]. Likewise, Noor *et al.* found moderately strong correlations between NTL features and a composite measure of wealth for several African regions [67]. However, they also reported that the strength of the correlations tended to reduce over time due to the penetration of electricity associated with development.

2.3.4 History

Other factors that influence aspects of liveability in cities are related to past historic events, political decisions, and trends. British cities underwent a process of unprecedented growth during the realm of Queen Victoria (1819 - 1901), the so-called "Victorian era". London population, for example, grew from 959,000 by 1800 to 2.76 million by 1860 [68]. The root causes of this growth were almost certainly associated with the hegemonic role of the British Empire in worldwide politics and commerce [69]. This accelerated even further when the Industrial Revolution gained momentum at the end of the 19th century. With the establishment of industries in urban areas, British cities grew at an even faster pace. Entire new neighbourhoods were built outside historic cores and the transportation of people from these to centres was made possible through the first railway lines [70]. Examples of extensive Victorian developments are the famous Liverpool's "Three Graces", a complex of several large buildings lining up on the seafront. These include the Mersey Docks and Harbour Board, Royal Liver and Cunard buildings. Although new developments were built, the migration of population from the poverty-stricken country to the recently industrialised cities was so conspicuous that the new homes were not enough and overcrowding became an issue of concern [71].

In the first half of the 20th century, World War II decisively affected the form of British cities and the distribution of their populations. All major urban centres

were, in fact, heavily bombed in what was called “the Blitz”, a series of night air raids carried out by Nazi Germany between 1940 and 1941. The targets of these attacks were not only industries and ports, but also civilian centres. The cities most affected by such bombing were London, Liverpool, Birmingham, and Glasgow [72].

More recently, in post-war UK, the joint effect of housing policies and design trends greatly influenced the distribution of socio-economic levels in cities. In the 1950s and 1960s, less advantaged people could not afford private housing and were thus placed in local authority housing which were designed in the style of the period (i.e., modernist style). This consisted of tower blocks retracted from the side walks, surrounded by open space and conspicuous presence of cul-de-sac. Conversely, better off people could pay for private higher quality historic housing (e.g., Georgian terraces), which tended to have the features of the traditional compact city form (i.e., well-connected street network, perimeter blocks). More detailed information on the relationship between post-war UK building policies, architectural styles, and socio-economic dynamics can be found, for example, in the works by Short [73] and Bullock [74].

Chapter 3

A Quantitative Method to Study Urban Form and City Liveability

In recent times, large repositories of geographic data have become openly accessible. This phenomena, which is commonly referred to as *open data revolution* [14, 15], has made available information on a wide range of topics. Among these, one can find data on the built environment but also on liveability indexes worldwide. In this thesis, I take advantage of this trend and develop a quantitative method that uses this data to quantify the relationship between urban form and city liveability.

The method works as follows. The first step consists in identifying past theories that discussed the relationship between urban form and liveability (box A in Figure 3.1). Jane Jacobs, for example, identified density, perimeter blocks, walkability, mix of different uses, mix of buildings of different ages and state of repair as linked to higher liveability [3]. Conversely, Le Corbusier despised the compact traditional city form and was in favour of a less dense, predominantly car-oriented, urban environment and tower blocks [5]. The second step involves the creation of metrics, inspired by these theories, that can function as proxies of urban form (box B in Figure 3.1). These metrics can be extracted from openly accessible datasets that contain information on the built environment (e.g., amenities, street network), such as OSM¹ or Foursquare² (box C in Figure 3.1). For example, given that both Jacobs and Le Corbusier considered density an important feature related to the liveability

¹<http://www.openstreetmap.org/>

²<https://developer.foursquare.com/>

of cities, one can translate this aspect in a quantitative metric. The third step consists in computing the metrics of urban form (box D in Figure 3.1). Since several metrics can be defined from the same theory, the fourth step consists in understanding what are the metrics that are best associated with a liveability indicator (box E in Figure 3.1). This is carried out by performing correlation analysis between each metric of urban form and a liveability index (box F in Figure 3.1). The relevant metrics are then used to build a comprehensive model, whose aims are to capture urban form as described by its multiple aspects and shed light on the relative impact of each metric in explaining the variance of liveability, in different areas (box G in Figure 3.1). The final step (arrow H in Figure 3.1) consists in linking the quantitative outcomes to past urban theories, to support findings, validate theories, or suggest changes.

In the remainder of this chapter I provide more details for each step.

3.1 Extraction of Metrics

The qualitative works presented in the previous chapter can provide inspiration for the definition of metrics, which capture different physical aspects of urban form. These works do not directly provide formulae; however, they clearly point to specific aspects of the urban environment, which can be translated into quantitative measures. Jacobs, for example, wrote about density, small blocks, walk-ability, mix of different uses, mix of buildings of different ages and state of repair [3]. Similarly, Whyte focused on human scale streets but also onto smaller details such as shop windows, porticoes, door steps [32]. Gehl mentioned the importance of active block frontages (i.e., city edges) [4]. Jacobs also focused on green areas and urban voids (e.g., big mono-functional buildings, wide highways) and how they affected the liveability of places, depending on their size and different urban contexts (e.g., densely built or not built at all) [3].

Operationally, these attributes refer to physical aspects of the built environment, such as size of blocks, connectivity of the street network, accessibility of places, but also to population density, built density, size and distribution of greenery. Furthermore, they also refer to amenities, uses, and activities at the ground

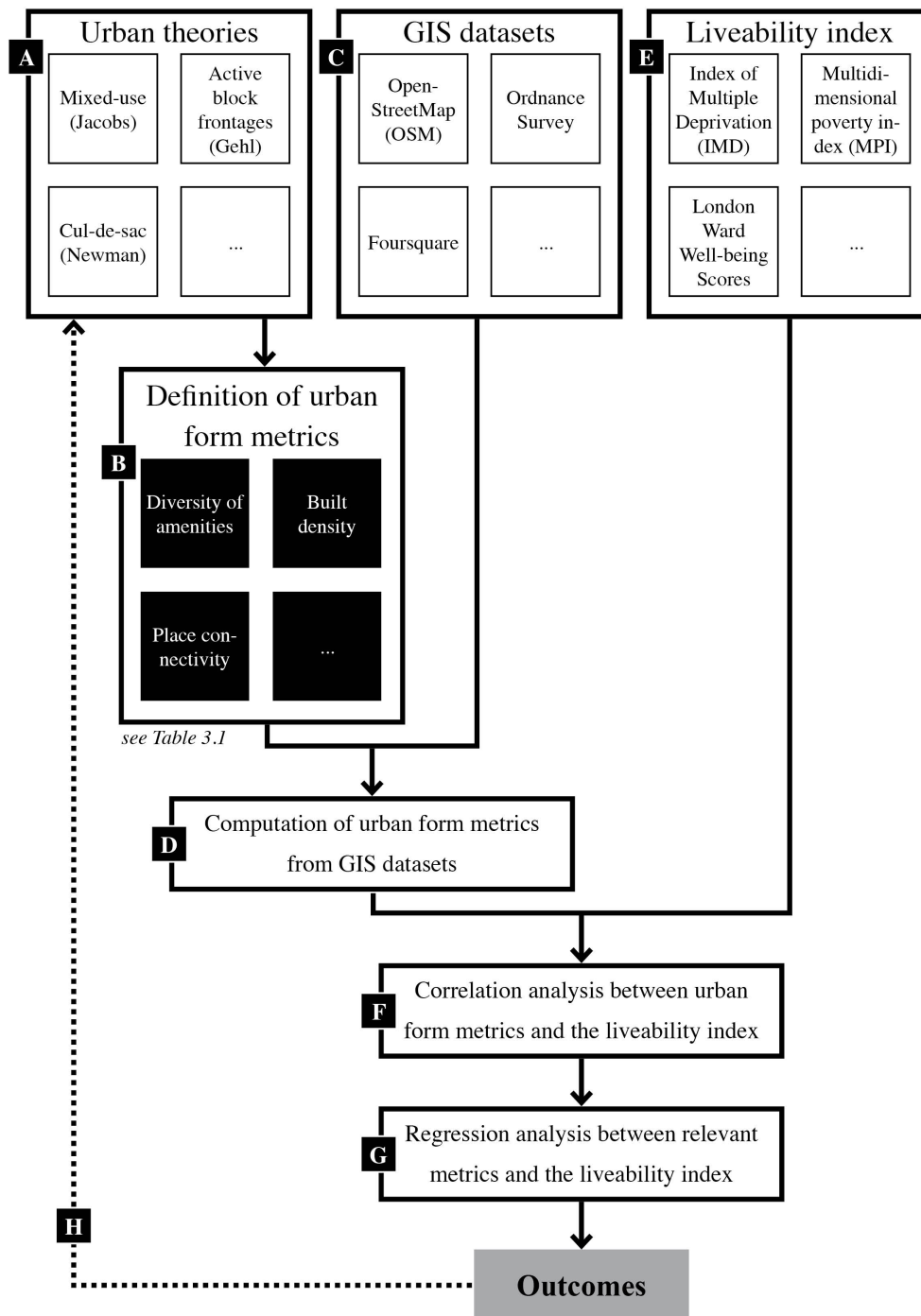


Figure 3.1: Flow chart of the methodology.

floor of buildings, for example, the presence of specific amenities and diversity of uses and buildings. Some examples of how qualitative urban theories can be trans-

formed into quantitative metrics is offered in Table 3.1 (more detailed formulae will be presented in the next chapters).

Generally, metrics of urban form capture two types of information and are extracted from different datasets:

Amenities. Measures of amenities and uses can be extracted from digital datasets containing information on their geographic locations (i.e., latitude and longitude). Foursquare is an example of this kind of datasets. The amenities and uses contained in Foursquare are very varied. They can be restaurants (e.g., Chinese restaurant, fast food restaurant), public buildings (e.g., city hall, courthouse), but also health services (e.g., dentist's office, hospital) and transport hubs (e.g., airport, bus station).

Configuration of the urban environment. Measures of the configuration of the urban environment (e.g., connectivity, block size, amount of greenery) can be extracted from digital datasets containing the representations of basic geographic features such as roads and intersections. OSM and Ordnance Survey (OS) VectorMap District³ are both examples of such datasets. From the former, one can obtain street intersections, which can be used to compute metrics of connectivity. From the latter, one can extract information on footprints of buildings. Such data is useful to calculate, for example, the occupancy ratio of land. From OSM, one can extract green areas to calculate amount of greenery.

Table 3.1: Examples of transformations of qualitative theories into quantitative metrics.

<i>Theory</i>	<i>Name of metric</i>	<i>Formula</i>
Well-connected streets promote social interactions, economic activities, and safety [3]	Place connectivity	n^o of intersections that are not cul-de-sac / total n^o of intersections
Medium to high densities stimulate urban life [4]	Population density	n^o of residents / area
Cul-de-sac create safer urban environments [41]	Density of cul-de-sac	n^o of cul-de-sac / area

³<https://www.ordnancesurvey.co.uk/business-and-government/products/vectormap-district.html>

3.2 City Liveability Indexes

As I presented in Chapter 2, there exist several indexes of liveability. These are usually based on the assumption that liveability is a multifaceted concept and are thus computed by considering several aspects of peoples' lives, such as income, education, health, among many others. Many of these indexes are computed at country level, for example, the *Where-to-Be-Born* [21] or the *Quality of Living* [22] indexes. However, the method proposed in this thesis aims to analyse a finer level of spatial granularity. Indexes that fit this requirement are, for example, the UK *Index of Multiple Deprivation* (IMD) [27], which is computed for small census areas of around 1,500 residents by considering seven different domains (such as income, education, crime among others); the *Multidimensional Poverty Index* (MPI) [26], which is provided at household level for over 100 developing countries and computed by considering health, education, and standard of living; the *London Ward Well-Being Scores* [75], which is calculated for the electoral districts of the London metropolitan region by considering eleven sub-domains (such as life expectancy, unemployment rate, and safety among others). These indexes are usually openly accessible and can be acquired through institutional websites, in the form of databases. A sample of the IMD dataset can be found in Appendix A.

3.3 Analytical Approach

Having identified the datasets from which to capture theory-based metrics and liveability indexes, the approach consists of four steps: (i) identifying the spatial unit of analysis and calculating the theory-based metrics and liveability for such unit; (ii) performing correlation analysis between each metric and liveability, to gain confidence that the theory-based metrics are actually reflecting the theories (and eventually to discard those that are not); (iii) performing regression analysis between relevant metrics and liveability, to understand to what extent urban form, described by multiple components, can explain of the variance of liveability; (iv) interpreting the outcomes of such analysis in light of previous urban theories and findings from other disciplines that study liveability in cities (e.g., Geography, Sociology). Next,

I present these steps in more detail.

3.3.1 Spatial Unit of Analysis

The spatial unit of analysis is the basic geographical entity for which the metrics are computed and aggregated. While there is no “single-size fits all” unit, several considerations should be taken into account when choosing it.

Firstly, analysing urban form means measuring the configuration of the street network, among other aspects. Metrics for quantifying this are usually based on the count of street intersections or node degrees (i.e., the count of the streets connected to an intersection). It is therefore important that the unit of analysis is big enough to allow the computation of such metrics. For example, a unit of analysis comprising a small 1 km by 1 km block size, would not be suitable for the analysis proposed in this thesis. Secondly, official spatial units, which existed for a long period of time, are better suited than more artificial grid-shaped units. Historical boundaries, in fact, usually provide the advantage of keeping the unity of neighbourhoods from a socio-cultural perspective as well as from a morphological one, for example by not cutting buildings or blocks. The former aspect is important when comparing the proposed metrics with measures of liveability which account for socio-economic aspects. The latter is important to have metrics of urban form which are consistent and as close as possible to the real features of neighbourhoods.

A further aspect that one should consider when selecting a spatial unit of analysis is the Modifiable Areal Unit Problem (MAUP) [76]. This can be a potential source of statistical bias which can affect outcomes of quantitative spatial analyses. Values of metrics can, in fact, vary substantially when computed for different boundaries. For example, population density, which is usually computed by dividing the number of residents by the extension of an area, can have different values if aggregated at the district level, for postcode areas, or for other spatial subdivisions. While there is no systematic approach to solve MAUP, one should acknowledge the issue and check whether it could constitute a source of bias for their study.

Once this selection is made, the next step consists in computing and aggregating the metrics of urban form and the liveability index for such unit. No aggregation

is needed for the liveability index if the spatial unit for which it is calculated is the same as the chosen one. However, especially in developed countries, liveability indexes are often provided at a small level of spatial granularity and thus may require aggregation. In developing countries, this may not be the case as indexes are often provided at a coarse level of granularity. In this case, the issue may actually be the opposite: the information is provided at a too coarse level.

3.3.2 Correlation, Regression Analysis, and Interpretations

The next steps consist in: (i) correlating the theory-based metrics with the selected index of liveability, to ascertain whether the former are good proxies for the theories; (ii) performing regression analysis between relevant metrics and liveability, to understand to what extent multiple features of urban form can explain of the variance of liveability; (iii) interpreting the quantitative outcomes of the analysis. To perform the first task, I use the Spearman correlation (often denoted as r_s) as it is robust in case data is skewed.

This technique measures the statistical dependence between two ranked variables, by assessing their monotonic relationship. It takes values between +1 and -1, where +1 corresponds to a perfect positive monotone relationship between two variables, 0 to a non existing relationship, and -1 to a perfect negative monotone relationship. The Spearman correlation is defined as the Pearson correlation between two ranked variables [77]. In mathematical terms, given a sample of size n , the n raw values X_i, Y_i are transformed into ranks rgX_i, rgY_i and r_s is computed as follows:

$$r_s = \rho_{rgX, rgY} = \frac{cov(rgX, rgY)}{\sigma_{rgX} \sigma_{rgY}}$$

where ρ corresponds to the Pearson correlation applied to the ranked variables, $cov(rgX, rgY)$ represents the covariance of the ranked variables, and σ_{rgX} and σ_{rgY} denote the standard deviations of the ranked variables. In the context of this thesis, X_i corresponds to each of the theory-based metrics being proposed and Y_i to a liveability index.

Since the method proposed is based on the interpretation of model outputs, I

use linear regression to perform the second task. This technique is the most suitable, among the many different types of regressions, as it affords a direct and easy reading of its outcomes. In linear regression, the model specification is that the dependent variable is a linear combination of a set of independent variables. A general linear regression for modelling n observations with p independent variables (i.e., the relevant theory-based metrics) takes the following mathematical form:

$$Y_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \varepsilon_i$$

where Y_i represents the values of the selected liveability index, i is the area under study, α represents the intercept. X_{i1} , X_{i2} , X_{ip} correspond to the relevant theory-based metrics; β_1 , β_2 , β_p are the parameters to be estimated, while ε_i represents the independent identically distributed normal error. The performance of a linear model is measured through the R^2 value. This quantifies the amount of variance in the dependent variable that is predictable from the independent variables. R^2 values tend to increase when more terms are used in a model. However, this increase in performance might not be due to an actual contribution of the added variables but just to the fact that the model has more terms. The adjusted R^2 value provides a better indication of model fit as it is adjusted for the number of terms used in the regression. This value only increases if the added terms improve the model more than it would be expected by chance. The adjusted R^2 value has thus to be preferred to the R^2 value for assessing the explanatory power of a linear model. Note that the linear regression technique presented in this paragraph can only be used under specific assumptions. For example, dependent and independent variables have to be normally distributed and linearly related. However, if these assumptions are not satisfied, dependent and independent variables can be transformed to meet them, for example through exponentiation. I discuss these methodological requirements in more detail, in the next chapters.

Finally, the outcomes of a linear regression analysis can be interpreted in the following ways. The adjusted R^2 value can be used to evaluate the overall capability of the model of explaining the variance of the liveability index, in different areas.

The β coefficients can be utilised to understand to what extent each theory-based metric is associated with liveability. This can be achieved by comparing the relative strengths and signs of these coefficients. The outcomes of this comparison can then be linked to previous urban theories, such as the ones by Jacobs [3] and Newman [41], and to results from other disciplines (e.g., Geography, Demography).

Note that this type of analysis does not imply causation. For example, even if model's outcomes suggest that more density is associated with better liveability, this does not necessarily mean that increasing density in specific neighbourhoods will correspond to an actual increase in liveability. Furthermore, findings are not generalizable as they are only valid for the specific geographic area and time frame for which the analysis is carried out. However, the methodology is replicable and can be used to perform the same analysis for other areas or for different time periods.

Chapter 4

On the Relationship Between Amenities and Socio-economic Deprivation

The first investigation of the proposed methodology focuses on the first hypothesis of this thesis, that is whether or not amenities are related to aspects of city liveability. More specifically, this investigation involves the analysis of the relationship between amenities and levels of socio-economic deprivation in three UK cities.

4.1 Introduction

As I presented in Chapter 2, previous works analysed the relationship between specific amenities in cities and aspects of city liveability, especially in the field of Health and Preventive Medicine. For example, several researchers found that health-promoting facilities were more concentrated in better-off neighbourhoods. Giles-Corti *et al.* found this to be true for golf courses in Australia [16], Powell *et al.* for fitness centres and dance facilities in the US [56]. Cummins *et al.* focused, instead, on a potentially health-harmful category of restaurants, that of fast food, and reported a strong relationship between their presence and socio-economically deprived neighbourhoods of England and Wales [17].

Inspired by this set of works, I apply the methodology proposed in this thesis to study the relationship between amenities in UK cities and levels of socio-economic

Table 4.1: Population, area, and density of the urban areas under study. Source: UK Census 2011.

<i>Urban area</i>	<i>Population</i>	<i>Area</i>	<i>Density</i>
Greater London	8,173,941	1,590 km^2	5,139 ppl per km^2
West Midlands	2,734,752	899 km^2	3,041 ppl per km^2
Greater Manchester	2,682,528	1,272 km^2	2,109 ppl per km^2

deprivation at a fine level of spatial granularity, through the use of openly accessible crowd-sourced geographic datasets (i.e., Foursquare and OpenStreetMap). First, I define metrics to capture what is present in city areas. Some amenities may uniformly cover a whole city, while others may be more or less concentrated in certain areas. It is this under/over representation of amenities that I capture in this investigation. I explain this in more detail later. Second, I identify what amenities are associated with socio-economic deprivation, through a correlation analysis. Third, I perform a regression analysis with these amenities as independent variables and socio-economic deprivation as dependent one. Finally, I interpret the outcomes of this analysis. I apply this approach to three UK urban regions of different size and population: Greater London, Greater Manchester, and West Midlands (the metropolitan region of Birmingham). Information regarding population count, density, and extension of these cities is presented in Table 4.1, while a map showing their geographic location in the UK is presented in Figure 4.1.

The remainder of this chapter is structured as follows. First, I illustrate the openly accessible datasets used to extract levels of socio-economic deprivation for the three cities under study and their relative amenities. Second, I present how I apply the general methodology presented in Chapter 3 to this specific analysis. Third, I use the method to analyse the relationship between the amenities of the three UK cities and levels of socio-economic deprivation. Finally, I report the outcomes, and provide interpretations before concluding with a final discussion.

4.2 Datasets

To carry out this analysis, I need information about the levels of socio-economic deprivation of the three cities under study and the amenities present in each of these cities. I extract the former from an official dataset containing the English Index

of Multiple Deprivation (IMD), while I obtain the latter from both Foursquare and OpenStreetMap (OSM). Note that IMD, Foursquare and OSM datasets are not temporally aligned as the first dates back to 2011, while the second two to 2014.¹ However, I assume that this would not affect the outcomes of this analysis as IMD has not changed significantly from its previous release (i.e., 2007). Indeed, variations of IMD scores, computed as differences between the 2007 scores and the 2011 ones, are small overall, with a mean variation of 0.7 and a standard deviation of 3.2. Note that the range of values of IMD in the 2007 and 2011 releases varies between a minimum value of 1.7 and a maximum value of 70.6. General information regarding the datasets used in the analysis presented in this chapter (i.e., IMD, Foursquare,

¹As of today, there exists newer versions of these datasets (e.g., IMD 2015). However, this study was conducted in 2014 and thus relied on datasets of that period.

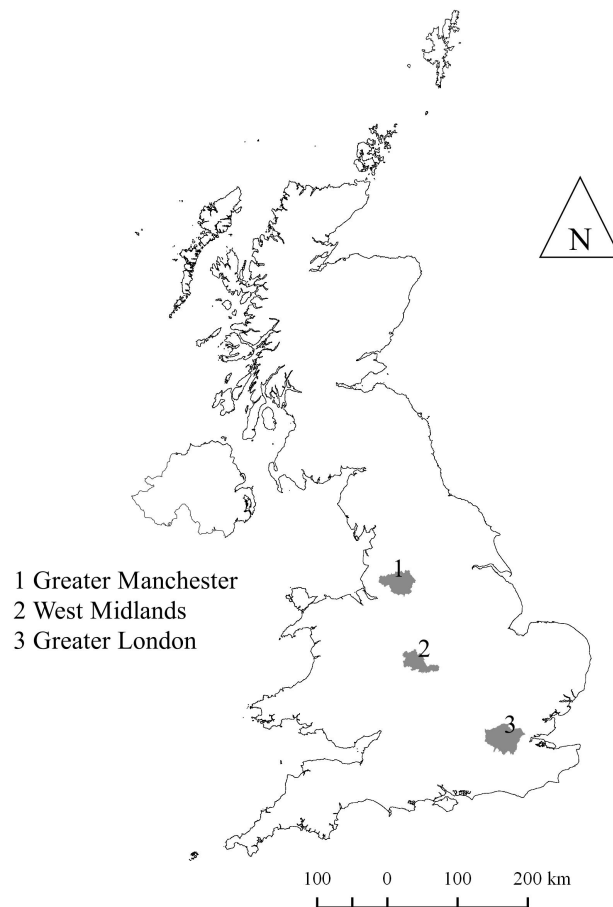


Figure 4.1: Location of the urban areas under study in the UK.

and OSM) is provided in Table 4.2, while more detailed descriptions can be found in the next subsections.

Table 4.2: General information on the datasets used in the analysis.

<i>Name</i>	<i>File type</i>	<i>Content</i>	<i>Spatial unit</i>	<i>Year</i>
Index of Multiple Deprivation (IMD)	Table	Composite measure of socio-economic deprivation	LSOA	2011
Foursquare	Table	Location of amenities and services	Geo-located point	2014
OpenStreetMap (OSM)	GIS dataset	Location of amenities and services	Geo-located point	2014

4.2.1 Index of Multiple Deprivation

As measure of socio-economic deprivation for the neighbourhoods of the three British cities under study, I use the UK official deprivation index (i.e., IMD) [27]. This index is issued by the UK Department for Communities and Local Government every four to five years and measures the relative deprivation of small census areas, called Lower-layer Super Output Areas (LSOAs),² by means of a ranking and a score. A low rank corresponds to a deprived area, while a high rank to a more advantaged one. Conversely, a high score corresponds to a deprived area, while a low score to a more advantaged one. As its name suggests, IMD is a composite measure, which means that is computed as a weighted mean of several different factors concerning people's living conditions. These are: income deprivation, employment deprivation, health deprivation, education deprivation, barrier to housing and services, crime levels, and living environment deprivation. Economic aspects (i.e., income and employment deprivation) are deemed of more relevance and thus have more weight in the calculation of the composite index. IMD generally follows a normal distribution [27]. Deprivation levels for the urban regions under study are taken from the 2011 survey. Figure 4.2 shows maps of deprivation for the three metropolitan regions under study (i.e., Greater Manchester, West Midlands, and Greater London).

²LSOAs are defined to roughly include the same number of residents (1,500) [27].

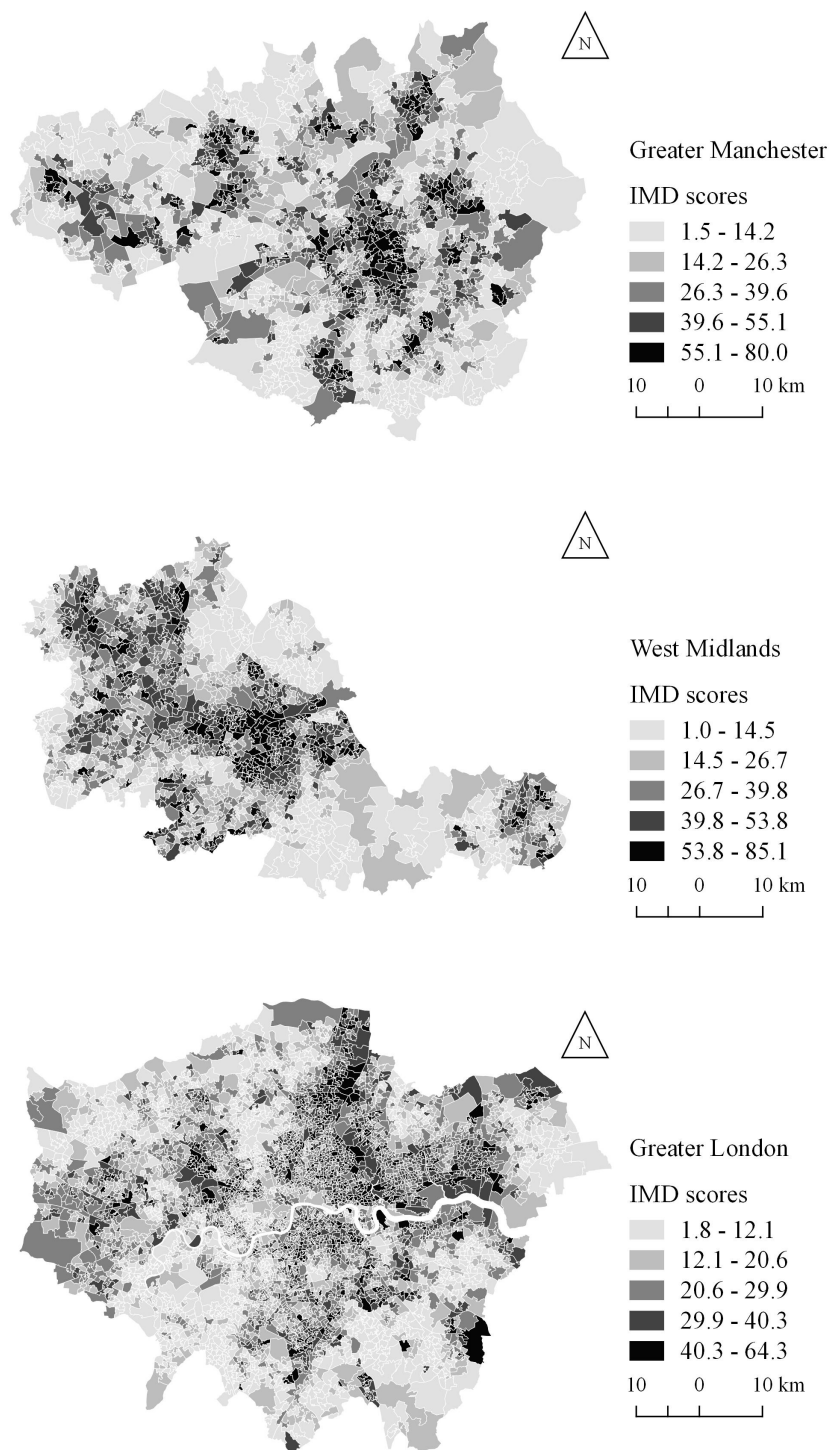


Figure 4.2: Maps of deprivation of the urban areas under study.



Figure 4.3: Foursquare amenities in central areas of the urban areas under study. Base map: OSM/Stamen.

4.2.2 Foursquare

As data source of amenities for the three urban regions under study, I use Foursquare, a mobile social-networking application launched in 2009 that enables people to register their presence in different locations through ‘check-ins’ and thus to share their whereabouts with virtual friends. It is reported that Foursquare is one of the most popular app for this purpose with more than 50 million people using its services each month.³ Beside checking-in, Foursquare users can also create new places. Possible conflicts or issues related to their definition are solved with a bottom up approach based on accuracy: the more accurate the description of a place is, the more likely it is that users are able to check-in into it. Multiple descriptions, which are likely to refer to the same place, are then merged by Foursquare through a Venue Harmonisation technique,⁴ which involves the use of developer-contributed geographic datasets. Each Foursquare place is defined by a set of geographic coordinates (i.e., latitude, longitude), a name, and a category (e.g., Chinese restaurant, University). Researchers have recently studied the reasons that make people checking-in and found several. However, one of these was particularly important to the creation of new places: people used Foursquare to remember visited places and curate their personal location history [78]. In contexts where Foursquare has high penetration (e.g., cities), the recorded places should thus constitute a well curated collection of existing amenities. Foursquare data for the three urban regions under study was obtained through its API⁵ in April 2014. I present a summary of the main features contained in the dataset in Table 4.3 and a map of Foursquare amenities for Central London in Figure 4.3.

<i>Urban area</i>	<i># Amenities</i>	<i># Check-ins</i>	<i># Categories</i>
Greater London	178,756	26,344,132	503
Greater Manchester	43,874	3,235,174	421
West Midlands	37,370	2,424,546	435

Table 4.3: Number of Foursquare amenities, check-ins, and categories across the urban areas under study.

³<https://foursquare.com/about>

⁴<https://developer.foursquare.com/overview/mapping>

⁵<https://developer.foursquare.com/>

4.2.3 OpenStreetMap

As additional source of information on amenities, I use OpenStreetMap (OSM), probably the best known example of geographic crowd sourcing. Contributors of this project are collectively building and keeping updated the first free, openly accessible, and editable map of the entire world. Its accuracy has been tested for several geographic contexts [79, 80, 81, 82] and found to be good overall, especially in urban areas. The OSM project has an ever growing number of users, which currently stands around 3 million.⁶ In OSM, real world features are represented through three types of spatial objects: *nodes*, *ways*, and *relations*. Nodes usually correspond to various amenities such as a pub or a school. Ways represent road segments. Relations are used to define geographic or logical relationships between spatial objects (e.g., bodies of water, bus routes). Since the focus of this analysis is on amenities, only nodes are considered. Similarly to a Foursquare place, an OSM node is defined by two main attributes: a pair of geographic coordinates (i.e., latitude, longitude) and a *tag* that describes it. This latter element comprises two attributes: a *key* (k) and a *value* (v). The former is used to label a category (e.g., leisure), the latter to provide details of such category (e.g., park). Here is an example in XML format:

```
<node id='358802885' ... lat='34.0666' lon='-118.7342'>
    ...
    <tag k='leisure' v='park' />
    ...
</node>
```

Differently from Foursquare, tags are not assigned following a top-down taxonomy, but can be freely determined by contributors. OSM node data for the cities under study is obtained in May 2014, through Geofabrik,⁷ a web service that extracts, selects, and processes OSM data. I provide a count of nodes and categories in Table 4.4 and a map of OSM amenities in Central London in Figure 4.4. Note

⁶<http://wiki.openstreetmap.org/wiki/Stats>

⁷<http://www.geofabrik.de/index.html>



Figure 4.4: OSM amenities in central areas of the urban areas under study. Base map: OSM/Stamen.

that both Foursquare and OSM datasets have biases, which means that they might not always represent accurately what is present in the real world. If this were the case, possible statistically significant associations between off-map amenities and deprivation would go unnoticed. Next, I present how I investigated the general methodology proposed in the previous chapter to fit the purposes of this analysis.

<i>Urban area</i>	<i># Amenities</i>	<i># Categories</i>
Greater London	79,343	896
Greater Manchester	24,321	381
West Midlands	27,885	465

Table 4.4: Number of OSM amenities and categories across the urban areas under study.

4.3 Approach

4.3.1 Metrics

Most of the amenities that a city offers are present in almost all areas. For example, in UK cities, pubs are present in almost every neighbourhood; the same could be said for schools as they provide a service of primary importance. A simple count of these features might then not be a significant proxy to characterise city areas. I thus propose the use of a formula called *Offering Advantage* (OA), as it is able to quantify to what extent a city area offers more or less of a specific amenity compared to the overall offering of that amenity throughout the whole city. OA is borrowed from Economy where it is called Revealed Comparative Advantage (RCA) and is used to quantify whether a country exports more or less of the good i (as a share of its total exports) than the average country [83]. This is computed by implementing the following formula:

$$RCA_{c,i} = \frac{goods_{c,i}}{goods_c} \cdot \frac{world}{world_i}$$

where $goods_{c,i}$ represents the total amount of goods i exported by the country c ; $goods_c$ is the total amount of goods exported by the country c ; $world$ denotes the total amount of goods exported all around the world; $world_i$ represents the total amount of goods i exported all around the world.

In the context of this analysis, this formula has been adapted to reflect to what

extent a neighbourhood n_k offers more of a certain amenity a_i than the average city area. To be more specific:

$$OA(a_i, n_k) = \frac{\text{count}(a_i, n_k)}{\sum_{j=1}^N \text{count}(a_j, n_k)} \cdot \frac{\sum_{j=1}^N \text{count}(a_j)}{\text{count}(a_i)}$$

where $OA(a_i, n_k)$ denotes the OA of the amenity a_i in the area n_k ; $\text{count}(a_i, n_k)$ represents the total number of occurrences of the amenity a_i in the area n_k ; N is the total number of categories of amenities; finally, $\text{count}(a_i)$ is the total number of occurrences of the amenity a_i in the whole city. I apply this formula to the amenities extracted from both Foursquare and OSM, across the three cities under study.

4.3.2 Spatial Unit of Analysis

In the previous chapter, I explained that the unit of analysis for studying features of urban form should have morphological unity and possibly be a long standing one. In this analysis, data comes in two forms: geographic points (i.e., latitude and longitude), for Foursquare and OSM; and LSOAs, for IMD. In theory, one could thus operate at the level of LSOAs, by aggregating Foursquare and OSM amenities at this level of spatial granularity. However, by inspecting the areas for which IMD is computed, I found that many areas were too small to be meaningful for the metrics used in this analysis. For example, many did not even contain the most common amenities (e.g., grocery store) or did not contain amenities at all. Furthermore, LSOAs were only recently created through an algorithm that only accounted for population density and that did not consider the morphological unity of city areas. For these reasons, I select a spatial unit of analysis, the *ward*, which is generally bigger than a LSOA, that accounts for the morphological aspect under study, and has existed for a longer period of time.⁸ The wards are official UK administrative boundaries which have both electoral and ceremonial function. The average extension of these spatial units in Greater London is of 255 hectares. I identify 625 wards for Greater London, 215 for Greater Manchester, and 163 for West Midlands. The OA values for each category of amenity and IMD are thus

⁸<https://www.ordnancesurvey.co.uk/election-maps/>

computed at this level of spatial granularity.

Since IMD is originally computed for LSOAs, which are spatial units smaller than wards, I aggregate IMD values of the LSOAs within each ward by computing the average of such values. One may wonder whether this process leads to spatial inaccuracies or loss of data. I found that this was not the case as groups of LSOAs perfectly fit within the boundaries of wards and the variation of IMD values associated with LSOAs within each ward is very low. Indeed, their standard deviation values are always smaller than their averages.

4.3.3 Correlation Analysis

Offering Advantage (OA) is computed for all the hundreds of different Foursquare and OSM amenities that the neighbourhoods of the three cities under study offer. However, I expect that not all of them would be significantly related to socio-economic deprivation. For example, one category of amenity (e.g., school) may be present both in the most and in the least advantaged areas. To select only the amenities that are associated with socio-economic deprivation, I thus perform a correlation analysis between the OA values of all the different categories of amenities and IMD scores. This involves implementing a correlation technique that accounts for spatial autocorrelation to quantify the relationship between the OA values of each category of amenity and deprivation, and addressing the issue of multiple testing. I provide more details next.

Given the spatial nature of the data analysed, one has to implement a correlation technique that accounts for spatial autocorrelation. In practice, this phenomenon represents the tendency of observations located near one another in space to be correlated. If this special kind of dependency is present in data, traditional techniques of correlation analysis such as Pearson and Spearman cannot be used as they require the independence of the observations. A preliminary test performed both on the OA values of categories of amenities and IMD indeed showed the presence of this phenomenon. I thus use a renowned method, common among natural scientists [84], and introduced by Clifford *et al.* [85], to account for spatial autocorrelation in data. This technique addresses the redundant information often present

in geographic data (i.e., spatial autocorrelation) through the computation of a reduced effective sample size. All the correlations presented in the Result section of this chapter are computed through the implementation of this technique.

A further step is needed to address the issue of multiple testing. Performing simultaneous correlation tests of hundreds of OA values can, in fact, potentially increase the chance of wrongfully rejecting the null hypothesis for some of them and thus increasing the chance of obtaining false positive results. To address this issue, I implement a control technique, in use among researchers who work with the testing of large numbers of distinct variables (e.g., Genomics), called False Discovery Rate (FDR) [86]. This method analyses the distributions of the p -values of tested variables and outputs a q -value, varying between 0 and 1, for each of them. These values quantify the probability of having false positives in the data. For example, a q -value of 0.04 means that 4% of all the variables with q -values below this threshold are false positives. While there is no widely recognized q -value cut-off, Bennet *et al.* suggested that a q -value of 0.1 would provide sufficient false positive protection [87]. I thus adopt 0.1 as cut-off threshold in this study. The q -values reported in the Result section of this chapter are computed by applying this technique to each of the OA values found to bear significant correlations.

4.3.4 Regression Analysis

The next step requires fitting a linear regression model with the OA values of the amenities that showed relevant associations with IMD as independent variables and IMD as dependent one. This requires a four-step process: (i) normalising variables to meet the assumption of linear regression; (ii) scaling the normalised variables to obtain comparable results; (iii) addressing possible collinearity among variables; (iv) performing stepwise linear regression to avoid model over fit (there might be hundreds of OA values that can potentially be regressed against IMD). I provide more details for each step next.

First, normalisation is required as it is an assumption of linear regression that candidate variables are normally distributed. In case these are not, various transformations can be applied to normalise them (e.g., exponentiation, logarithmic trans-

formation). Second, scaling is necessary as the OA values may have different magnitudes and may thus be hard to compare and interpret, if regressed untransformed against IMD. Scaling consists in calculating the standard scores (i.e., z scores) of the normalised OA values. This can be done through the following formula:

$$z = \frac{X - \mu}{\sigma}$$

where X represents the raw value, μ the mean of the population, and σ its standard deviation.

The third step requires testing relevant OA values for collinearity. This phenomenon occurs when two or more candidate variables for a regression model are strongly correlated among each other. If strongly collinear variables are used in a regression, it is likely that the resulting fit and regression coefficients are inflated or show unexpected signs. To check for this issue, I compute the Variance Inflation Factors (VIFs) associated with each candidate variable. Let reg be a regression model with predictor variables v_1, v_i, \dots, v_n . The VIF of the variable v_i is obtained by, first, performing linear regression with v_i as dependent variable and the other variables as independent ones $v_1, v_i - 1, v_i + 1, \dots, v_n$, and, second, by using the overall model fit (i.e., R^2 value) obtained at the previous step in the following formula:

$$VIF = \frac{1}{1 - R^2}.$$

If a variable has a strong linear relation with at least another one, its correlation coefficient is likely to be close to 1 and the VIF related to that variable large. A VIF equal to or greater than 10 is a sign of a collinearity issue [88]. If the candidate variables do not show VIFs above 10, they can be used as independent variables in a linear regression model with IMD as dependent one. Conversely, if candidate variables have VIFs equal to or greater than 10, one can implement a stepwise procedure that, first, excludes the candidate variable with the highest VIF and then repeats the same process until none of the variable has a VIF equal to or greater than 10. At the end of this procedure, candidate variables should be devoid of collinearity

and can be regressed against IMD.

The final step requires performing stepwise linear regression between the relevant OA values and IMD. This technique is necessary as there might be hundreds of OA values that can potentially be regressed against IMD. A simple linear regression would output a model with too many coefficients that would be hard to interpret and plausibly run into the issue of over fitting. Conversely, a stepwise linear regression outputs a parsimonious model that relies on the smallest number possible of regression coefficients. Having obtained a model through this technique, a reliability check has to be performed on the outcomes of such model. In the previous section, I explained how spatial autocorrelation can bias the outcomes of a correlation analysis. The same phenomenon can also affect the outcome of a stepwise linear regression by causing over-inflated regression coefficients or unexpected signs. This is caused by not considering the information derived from the spatial dependency of observations located close to each other in space. For this reason, it is necessary to check whether spatial autocorrelation is present in the outcome of a regression too. To this end, I use a technique renowned in spatial studies, the Moran's test [89], that checks whether residuals are spatially autocorrelated. The outputs of this test are two. An index (I) that varies between -1 and 1, which can be interpreted similarly to a Pearson's correlation coefficient, and a p -value that measures the statistical significance of the test. A negative Moran's I generally occurs when the spatial tendency is such that dissimilar values cluster together (dispersed pattern). Conversely, it is positive when similar values are located near each other (clustered pattern). If the output of this test shows that there is no statistical evidence that spatial autocorrelation is present in the residuals, the outcome of the regression can be accepted and interpreted. Conversely, if this is not the case, one should consider regression techniques that incorporate the spatial information in their equations. One of these techniques – and the one adopted in this study – is the Spatial Autoregressive (SAR) model, a type of regression that accounts for the proximity of observations in space by including a spatial weighting matrix in the equation [90]. More specifically, the

general formula for a SAR model is:

$$Y_n = \lambda W_n Y_n + X_n \beta + \varepsilon_n,$$

where Y_n is the dependent variable; W_n is an n by n spatial weighting matrix⁹ applied to the variable Y_n , with λ being a spatial autoregressive parameter estimated from the data. X_n is the independent variable, β is the regression coefficient associated with X_n , and ε_n is the error term. In practice, this model expresses the concept that the value of a variable at a given location is associated with the values of the same variable measured at nearby locations, reflecting a sort of interaction effect. To ascertain that the SAR model accounts for all the spatial autocorrelation present in the data, I utilise again the Moran's test on the residuals of the SAR model. If the Moran's I is close to 0 or is not statistically significant (p -value > 0.05), the SAR model can be accepted and interpreted. Otherwise, it should be rejected.

I present in the next section the results of the analysis performed on the three UK urban areas under study (i.e., Greater London, Greater Manchester, and West Midlands).

4.4 Results

In this section, I first report the results of the correlation analysis used to identify what OA values of Foursquare and OSM categories of amenities, among hundreds, were associated with deprivation. Second, I present the outcomes of the three regression models (one for each urban area) with the relevant OA values as independent variables and IMD as dependent one. Finally, I interpret the quantitative results.

4.4.1 Amenities and Deprivation

To identify what Foursquare and OSM amenities are associated with deprivation, I first compute the *Offering Advantage* for each category of amenity, for each urban area, for both datasets. I then calculate the Spearman's rank correlation coefficient

⁹Where n corresponds to the number of observations (areas) considered.

(r_s) between the OA value of each category of amenity and deprivation, through the implementation of the Clifford *et al.* method. I present in Table 4.5 and Table 4.6 the number of Foursquare and OSM amenities correlated with IMD, grouped by strength of correlation.

<i>Urban area</i>	$ r_s \in [0.05, 0.2)$	$ r_s \in [0.2, 0.4)$	$ r_s \in [0.4, 0.6)$
Greater London	23	15	0
Greater Manchester	30	24	0
West Midlands	17	33	1

Table 4.5: Number of Foursquare categories of amenities correlated with IMD (all results shown are statistically significant, $p < 0.05$)

<i>Urban area</i>	$ r_s \in [0.05, 0.2)$	$ r_s \in [0.2, 0.4)$	$ r_s \in [0.4, 0.6)$
Greater London	4	2	0
Greater Manchester	1	4	0
West Midlands	1	9	0

Table 4.6: Number of OSM categories of amenities correlated with IMD (all results shown are statistically significant, $p < 0.05$)

By inspecting the tables, I highlight two remarks. First, there is a conspicuous amount of amenities (i.e., 166 across the three cities), extracted from both datasets, correlated with deprivation. To be more specific, 76 are weakly correlated, 87 are weakly to moderately correlated, and only one is moderately correlated with IMD. This suggests that there is a wealth of amenities associated with deprivation across urban areas of different size, density, and geographic location. The second observation concerns the different results obtained with Foursquare and OSM data. The number of Foursquare amenities correlated with deprivation across the three cities (i.e., 143), far exceeds the number of OSM ones (i.e., 21). It thus appears that the amenities being mapped in OSM, at the time of this study, are conceptually less related to deprivation in the three cities under study. I present in Table 4.7 and Table 4.8 the top three positively and negatively correlated Foursquare and OSM amenities. Results suggest that the two datasets provide two different types of information. On one side, Foursquare tends to offer more information on services and retail (e.g., Bus Station, Italian restaurant). On the other, OSM seems to provide more information on elements of the road system (e.g., traffic signals, crossing). The two datasets seem, thus, to complement one another.

	<i>Greater London</i>	<i>Greater Manchester</i>	<i>West Midlands</i>
Top positively correlated	Caribbean rest. (0.37) African rest. (0.32) fried chicken (0.31)	bus station (0.32) residential (0.27) student centre (0.24)	car wash (0.38) temple (0.34) desserts (0.32)
Top negatively correlated	Indian rest. (-0.27) Italian rest. (-0.26) golf course (-0.24)	Italian rest. (-0.36) golf course (-0.28) gastropub (-0.26)	golf course (-0.42) salon barbershop (-0.35) farm (-0.31)

Table 4.7: Foursquare categories of amenities most correlated with IMD. In parentheses, the Spearman r_s (statistical significance level: $p < 0.05$) computed with the Clifford *et al.* method.

	<i>Greater London</i>	<i>Greater Manchester</i>	<i>West Midlands</i>
Top positively correlated	traffic signals (0.29) crossing (0.25) community centre (0.18)	traffic signals (0.25) taxi (0.24)	tram stop (0.31) billboard (0.29) artwork (0.29)
Top negatively correlated	parking (-0.14) garden centre (-0.10)	post box (-0.26) kindergarten (-0.22) restaurant (-0.18)	parking (-0.30)

Table 4.8: OSM categories of amenities most correlated with IMD. In parentheses, the Spearman r_s (statistical significance level: $p < 0.05$) computed with the Clifford *et al.* method.

As I mentioned before, the multiple correlation tests of hundreds of variables can increase the chance of wrongly rejecting the null hypothesis for some of them. To address this issue, I implement the FDR technique by first ranking the variables according to their relative p -values, from low to high, and by calculating the q -values on the ranked list. I present in Table 4.9 a summary of these outputs, subdivided for cities and datasets. The values presented correspond to the q -values computed for each quartile of the ranked p -values. For what concerns the results obtained with Foursquare data, the amount of false positives is less than 22% across all amenities for Greater Manchester. This value raises slightly (i.e., 24%) when considering the West Midlands case. For Greater London, the amount of false positives is at most 17% for half of the variables (those with the lowest p -values). However, it raises to 47% when considering the whole set. For what regards the outcomes obtained with OSM data, the amount of false positives is generally quite high. For Greater London and Greater Manchester, up to 85% and up to 51% of the correlations are false positives, respectively. The only exception is West Mid-

lands. For this case, q -values are lower across the whole set of variables with the amount of false positives being at most 28%. These results seem to suggest that the Foursquare dataset is more suited than OSM for the purpose of this analysis.

<i>Dataset</i>	<i>Urban area (# categories)</i>	<i>1st Qu.</i>	<i>Median</i>	<i>3rd Qu.</i>	<i>Max.</i>
Foursquare	Greater London (35)	0.06	0.17	0.36	0.47
	Greater Manchester (54)	0.04	0.07	0.12	0.22
	West Midlands (50)	0.07	0.12	0.16	0.24
OSM	Greater London (6)	0.44	0.44	0.75	0.85
	Greater Manchester (5)	0.31	0.39	0.50	0.51
	West Midlands (10)	0.13	0.16	0.26	0.28

Table 4.9: Q -values for quartiles of p -values, computed through the False Discovery Rate technique.

4.4.2 Modelling Amenities and Deprivation

Having identified the categories of amenities that are mostly correlated with the deprivation levels of the three cities under study, I proceed to build a comprehensive model by using all the OA values, with q -values smaller than 0.1, as independent variables and IMD as dependent one. To this end, I first normalise and scale such variables, as the majority showed skewed distributions. Second, I test whether they show collinearity through the Variance Inflation Factor (VIF) test. Outcomes of this test show that none of the OA values present the issue, in none of the cases. Greater London has thus 11 candidate variables, Greater Manchester 31, and West Midlands 22. I then perform three separate stepwise linear regressions (one for each urban area) with the OA values as independent variables and IMD as dependent one. Outcomes show that the selected categories of amenities can explain, at 99% confidence level, 34% of the variance of IMD in Greater London, 51% in Greater Manchester, and 56% in West Midlands. However, the Moran's test finds a statistically significant presence of spatial autocorrelation in the residuals of the three models. The Moran's I in the model for Greater London is 0.44, it is 0.12 in the one for Greater Manchester, and 0.22 in the one for West Midlands. I thus utilise the SAR model, which accounts for spatial autocorrelation, to regress the variables selected by the stepwise procedure previously implemented against IMD. Outcomes indeed show that part of IMD is explained by the spatial factor. Incorporating this

information in the equation causes, in fact, an increase of the adjusted R^2 values and a slightly decrease of the regression coefficients. Nonetheless, almost all of the latter maintain statistical significance. The Moran's test performed on the residuals of the SAR models shows that there is no statistical evidence of the presence of spatial autocorrelation (all p -values are far greater than 0.05). The SAR model for Greater London can explain 73% of the variance of IMD, the one for Greater Manchester 61%, while the one for West Midlands 75%. In the model for the first urban area, all 10 regression coefficients are statistically significant, at 90% confidence level, in the one for the second urban area, 14 out of 21 are significant, in the one for the third, 14 out of 17 are significant. The two strongest regression coefficients in Greater London are *fried chicken* ($\beta = 0.11$) and *bank* ($\beta = -0.09$), in Greater Manchester are *bus station* ($\beta = 0.17$) and *Italian restaurant* ($\beta = -0.16$), while in West Midlands are *salon barbershop* ($\beta = -0.15$) and *desserts* ($\beta = 0.14$). Please refer to Table 4.10, 4.11, 4.12, for more details on the outcomes of the linear regression (LR) and SAR models for Greater London, Greater Manchester, and West Midlands, respectively.

Table 4.10: LR and SAR models for Greater London. Red bar means positively associated with IMD, blue bar means negatively associated with IMD. '.' significant at $p < 0.1$; '*' significant at $p < 0.05$; '**' significant at $p < 0.01$; '***' significant at $p < 0.001$. The Foursquare categories can explain up to 73% of the variance of IMD, with the strongest coefficients being *fried chicken* (i.e., $\beta = 0.11$) and *bank* (i.e., $\beta = -0.09$).































































Independent variable	LR		SAR	
	p -value	β	p -value	β
(intercept)		0.00		-0.01
<i>African restaurant</i>	***	0.20 	***	0.08 
<i>bank</i>	***	-0.12 	***	-0.09 
<i>cricket</i>	**	-0.10 	**	-0.07 
<i>dentist</i>	***	-0.17 	***	-0.08 
<i>factory</i>	***	0.14 	***	0.07 
<i>fried chicken</i>	***	0.20 	***	0.11 
<i>golf course</i>	***	-0.16 	***	-0.06 
<i>grocery store</i>	***	0.17 	***	0.08 
<i>mosque</i>	***	0.17 	***	0.08 
<i>salon barbershop</i>	**	-0.11 	**	-0.06 
Adjusted R^2		0.34		0.73
p -value		0.00		0.00
Moran's I		0.44		-0.07
p -value		0.00		1.00

Table 4.11: LR and SAR models for Greater Manchester. Red bar means positively associated with IMD, blue bar means negatively associated with IMD. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘**’ significant at $p < 0.01$; ‘***’ significant at $p < 0.001$. The Foursquare categories can explain up to 61% of the variance of IMD, with the strongest coefficients being *bus station* (i.e., $\beta = 0.17$) and *Italian restaurant* (i.e., $\beta = -0.16$).

<i>Independent variable</i>	<i>LR</i>		<i>SAR</i>	
	<i>p</i> -value	β	<i>p</i> -value	β
(intercept)		0.00		-0.02
<i>administrative building</i>	*	0.11 		0.07 
<i>bus station</i>	**	0.17 	***	0.17 
<i>camp ground</i>		-0.09 	.	-0.08 
<i>car wash</i>	.	0.10 	*	0.10 
<i>community college</i>	.	0.10 		0.06 
<i>dentist</i>	.	-0.09 		-0.07 
<i>fast food</i>	*	0.12 	**	0.12 
<i>field</i>	**	-0.14 	**	-0.12 
<i>flower shop</i>	**	-0.14 	**	-0.13 
<i>gastropub</i>	.	-0.09 		-0.07 
<i>gas station garage</i>	*	0.11 	**	0.12 
<i>golf course</i>	.	-0.10 	.	-0.08 
<i>government</i>		0.07 		0.06 
<i>Italian restaurant</i>	***	-0.18 	***	-0.16 
<i>other outdoors</i>	*	-0.10 	*	-0.10 
<i>playground</i>	*	-0.11 	*	-0.10 
<i>radio station</i>		0.08 		0.07 
<i>student centre</i>	***	0.17 	**	0.15 
<i>supermarket</i>	*	0.10 	*	0.09 
<i>tennis court</i>		-0.08 		-0.07 
<i>trail</i>	.	-0.10 	.	-0.09 
Adjusted R^2		0.51		0.61
<i>p</i> -value		0.00		0.00
Moran's I		0.12		-0.02
<i>p</i> -value		0.00		0.61

4.4.3 Interpretations

In the previous section, I illustrated how different sets of categories of amenities can explain the deprivation of the three UK cities under study. However, the relative large number of categories for each set makes it hard to interpret the outcomes. I thus use an inductive thematic analysis [91] on the OA values of the amenities in the models. This procedure requires three steps: two should be separately conducted by two or more people, while the last should be carried out together. The first step requires reading through the Foursquare and OSM categories of amenities and creating linguistic codes for each of them. The second step involves merging codes that are semantically and conceptually related to broader topics. The last step, which

Table 4.12: LR and SAR models for West Midlands. Red bar means positively associated with IMD, blue bar means negatively associated with IMD. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘**’ significant at $p < 0.01$; ‘***’ significant at $p < 0.001$. The Foursquare categories can explain up to 75% of the variance of IMD, with the strongest coefficients being *salon barbershop* (i.e., $\beta = -0.15$) and *desserts* (i.e., $\beta = 0.14$).

<i>Independent variable</i>	<i>LR</i>		<i>SAR</i>	
	<i>p-value</i>	β	<i>p-value</i>	β
(intercept)		0.00		-0.04
<i>African restaurant</i>	*	0.15	*	0.09
<i>breakfast</i>	.	0.11	*	0.09
<i>car wash</i>		0.09		0.03
<i>community college</i>	.	0.11	**	0.13
<i>courthouse</i>	**	0.15	**	0.12
<i>cricket</i>	*	-0.13	*	-0.11
<i>desserts</i>	***	0.19	**	0.14
<i>farm</i>	*	-0.13	.	-0.07
<i>flower shop</i>	*	-0.14	*	-0.11
<i>golf course</i>	.	-0.11	.	-0.09
<i>Italian restaurant</i>	***	-0.19	*	-0.09
<i>laundry</i>	.	-0.10	.	-0.08
<i>light rail</i>	.	0.11		0.02
<i>other outdoors</i>	*	-0.14	**	-0.12
<i>salon barbershop</i>	**	-0.16	***	-0.15
<i>supermarket</i>	*	0.12	.	0.07
<i>temple</i>	.	0.10		0.05
Adjusted R^2		0.56		0.75
<i>p-value</i>		0.00		0.00
Moran's I		0.22		-0.02
<i>p-value</i>		0.00		0.59

should be carried out by all the people who conducted the exercise, consists in refining and agreeing the names of topics. The outcome of this process consists thus of a set of topics that represent simple and distinctive urban characteristics associated with deprivation, in the three cities under study.

The thematic analysis for the categories identified in the regressions is carried out by the author of this thesis, who has a background in Urban Design, and by Giovanni Quattrone, a colleague with a background in Computer Science. We separately went through the categories and we associated with each of them some terms that we thought described them well. For example, I associated *heavy* and *unhealthy* with the Foursquare category *African restaurant*, while Giovanni chose *ethnic* and *greasy*. We then separately analysed the terms identified at the previous step and grouped those that were semantically and conceptually related in broader

topics. For example, I created the topic *fat food* to include the categories *African restaurant*, *fried chicken*, and *fast food*. Giovanni generated the topic *sports* to include the categories *cricket*, *golf course*, and *tennis court*. At this point, both of us got together and merged the topics previously identified into the final ones. For example, my topic *fat food* was merged with Giovanni's *unhealthy restaurants* to create the final topic *unhealthy food*. This is because some of the categories that offer unhealthy meals, considered in both our topics, cannot be described as restaurants (i.e., breakfast, desserts). At the end of this process, we identified a final set of five topics: *surplus goods*, *sport & recreation*, *unhealthy food*, *unpleasant places*, and *public buildings*.

Surplus goods. This topic brings together those categories of amenities which offer services or goods that are not strictly necessary for living (i.e., surplus goods). These are *dentist*, *salon barbershop*, and *flower shop*. These categories bear a negative relationship with IMD, meaning that they tend to concentrate in more advantaged neighbourhoods. This might be due to the fact that those areas might have a population who is more willing to pay for those goods. Findings from previous literature seem to corroborate this hypothesis, at least for the category *dentist*. Locker reported, in fact, that an higher socio-economic status was linked to better oral health [92].

Sport & recreation. This topic comprises those categories of amenities that are linked to physical activity, leisure time, and recreation. These are *cricket*, *golf course*, *camp ground*, *field*, *other outdoors*, *playground*, and *trail*. These categories are inversely related to deprivation, meaning that they tend to concentrate in more advantaged neighbourhoods. I hypothesise that the reason behind this relationship lies in the fact that better-off people tend to have more time and resources to dedicate to such activities. This finding seems to find support in previous works that reported that golf courses in Australia [16] and fitness and dance facilities in the US [56] tend to be more concentrated in more advantaged neighbourhoods. Moreover, other works reported that the presence of open spaces was strongly associated with better-off city areas, in the Netherlands [93], in Howard County, US [94], and in

Portland, US [95].

Unhealthy food. This theme includes those categories of amenities that provide unhealthy food. These are *African restaurant, fried chicken, fast food, breakfast, and desserts*. These categories are positively associated with deprivation, meaning that they tend to be located in more deprived areas. It is possible that the consumption of unhealthy food (e.g., greasy, fried, high in sugar) is associated with the eating habits of more deprived population. This hypothesis seems to be corroborated by more than one study, at least for the category *fast food*. MacDonald *et al.* found that higher density of fast-food chain restaurants corresponded to higher deprivation in England and Scotland [17]. Pearce *et al.* and Block *et al.* reported similar outcomes in New Zealand [96] and in the US [97], suggesting that the relationship is true for more than one country.

Unpleasant places. This topic brings together those categories of amenities that are usually surrounded by an unpleasant urban environment, characterised, for example, by big parking lots, highways, dangerous intersections, and air pollution. These are *factory, bus station, car wash, and gas station garage*. These categories are positively associated with IMD, meaning that they are located in more deprived neighbourhoods. It is likely that the presence of these amenities, with their relative surroundings, deteriorate the well-being of city dwellers. This seems to find support in one study, at least for the category *factory*. Perlin *et al.* reported a significant link between poor households, in three different US states, and the presence of one or more industrial sources of air pollution nearby [98].

Public buildings. This theme comprises those categories of amenities that provide certain public functions. These are *student centre, community college, courthouse*. These categories are associated with deprivation, meaning that they tend to be present in more deprived neighbourhoods. I suggest that this relationship might be explained by the fact that less advantaged residents might be more in need of the services provided by those amenities. However, I did not find literature to support this hypothesis.

4.5 Discussion

In this chapter, I tested the first hypothesis of this thesis (i.e., amenities are associated with aspects of city liveability) to study the relationship between the amenities extracted from openly accessible datasets (i.e., Foursquare and OSM) and deprivation. Outcomes suggest that this link exists and that sets of categories of amenities explain some variance of the deprivation index, across the cities considered in this study. Common patterns of categories of amenities related to deprivation have been discovered, across the three cities, through thematic analysis. To be more specific, *surplus goods* (e.g., dentist, salon barbershop) and *sport & recreation* (e.g., golf course, field) were found to be associated with more advantaged neighbourhoods. Conversely, *unhealthy food* (e.g., fast food, fried chicken), *unpleasant places* (e.g., factory, bus station), and *public buildings* were found to be related to more deprived neighbourhoods. I follow next with the implications and limitations of this analysis.

4.5.1 Limitations

I ought to acknowledge some limitations for the study presented in this chapter. First, both Foursquare and OSM have geographic and social biases. They do not have a uniform coverage across space. They tend to provide more information in city centres and less in peripheries, thus offering only a partial picture of what is actually present in the real world [99]. For what concern the social bias, users of both Foursquare and OSM tend to belong to the same social group (i.e., young, educated, and wealthy) [99] and thus one may question whether the content of these datasets corresponds to what is actually present in the real world. When applying the proposed approach, one should first check whether the area that she or he wants to analyse is subject to geographical or social bias, for example by using the method proposed by Quattrone *et al.* [100]. If the biases are large, it is likely that the results obtained through the proposed approach are unreliable.

A second limitation is associated with the use of OSM data. Since OSM users can freely tag map elements (i.e., the tagging system does not follow a pre-established taxonomy), it is possible that ambiguous denominations are assigned to amenities. This might introduce inaccuracies in the computation of the OA values

and thus in the outcomes of the analysis.

A third limitation concerns the multiple correlation testing. The proposed approach requires the simultaneous testing of hundreds of variables. This may lead to the erroneous exclusion of the null hypothesis for some of the correlations and thus to an increased chance of obtaining false positives. In this analysis, this threat is estimated through the FDR technique. However, it does not solve the problem completely. When applying the proposed approach, one should thus estimate the risk of invalid findings through the FDR technique and make a decision based on this estimation to whether or not accept the results and continue with the analysis. For the specific case of the three cities under study, results seemed robust especially for Foursquare data and for the urban areas of Greater Manchester and West Midlands.

A fourth limitation concerns the presence of unexpected regression coefficients in the models. Since the filtering procedure proposed is automatic, it is possible that the relationships between some categories of amenities and deprivation are hardly explainable. This is the case, for example, of *grocery store*, associated with deprivation, in Greater London, and of *laundry*, related to more advantaged neighbourhoods, in West Midlands.

A further limitation regards generalizability. The outcomes of this analysis only hold for the three cities under study (i.e., Greater London, Greater Manchester, and West Midlands) and for 2011. It is thus impossible to derive universal findings from such outcomes. Nonetheless, the very same approach presented in this chapter can be applied to different geographic contexts and time frames to ascertain whether findings hold or not.

A final limitation regards the direction of causality. The approach presented in this chapter has to be considered an exploratory technique, which can be applied to test research questions concerning the presence of specific amenities in neighbourhoods and deprivation. The outputs of this analysis, in fact, do not establish the direction of causality between specific amenities and socio-economic deprivation. This means, for example, that, although fried chicken restaurants might be found to be related to deprivation, this does not necessarily mean that decreasing the number

of this category of restaurants would correspond to an actual improvement of the socio-economic levels of city dwellers.

Chapter 5

On the Configuration of the Urban Environment and Socio-economic Deprivation

In the previous chapter, I illustrated how sets of specific categories of amenities were associated with different levels of deprivation (i.e., the first hypothesis of this thesis) in three UK urban areas. In this chapter, I test the second hypothesis of this thesis (i.e., configurational aspects of the urban environment are associated with aspects of city liveability), by applying the proposed methodology to the deprivation levels of six UK urban areas.

5.1 Introduction

As I illustrated in Chapter 2, different urban theories have been developed in the twentieth century. On one side, there were those supporting the traditional compact city form, such as Jane Jacobs. On the other, there were those favouring more spread out and car-oriented urban developments, such as Le Corbusier. More recently, researchers have adopted quantitative methods for studying the relationship between cities and aspects of well-being. However, both qualitative and quantitative works have limitations. The former are hardly replicable and generalizable, while the latter tend to study single aspects of the urban environment in relation to aspects of liveability and thus fail to present a more complete picture of what kind

Table 5.1: Population, area, and density for the six urban areas under study. Source: UK Census 2011.

<i>Urban area</i>	<i>Population</i>	<i>Area</i>	<i>Density</i>
Greater London	8,173,941	1,590 km^2	5,139 ppl per km^2
West Midlands	2,734,752	899 km^2	3,041 ppl per km^2
Greater Manchester	2,682,528	1,272 km^2	2,109 ppl per km^2
Leeds	1,723,770	940 km^2	1,834 ppl per km^2
Liverpool	1,245,929	521 km^2	2,391 ppl per km^2
Newcastle	1,145,353	577 km^2	1,986 ppl per km^2

of neighbourhoods is related to such aspects. In this analysis, I apply the methodology presented in Chapter 3 to study the relationship between a set of measures of urban form and the levels of socio-economic deprivation of six UK urban areas: Greater London, Greater Manchester, West Midlands, Liverpool, Leeds, and Newcastle. To carry out the analysis, I extract metrics of the urban environment from Ordnance Survey (OS) VectorMap District and OpenStreetMap (OSM), while I gather socio-economic data of neighbourhoods from the 2011 Index of Multiple Deprivation (IMD). I provide information concerning population count, extension, and density of the six urban areas under study in Table 5.1, while I present a map of their locations in the UK in Figure 5.1.

The remainder of this chapter is structured as follows. I first illustrate what datasets are needed to carry out the analysis. Second, I present the quantitative approach, based on the methodology presented in Chapter 3. Finally, I illustrate the outcomes of the analysis and provide interpretations.

5.2 Datasets

Since this analysis focuses on the configuration of urban form and socio-economic deprivation, two types of datasets are necessary: one containing representations of real word features (e.g., roads, buildings) and another one providing levels of deprivation for the urban areas under study. OS VectorMap District and OSM are used as data sources for the former elements, while the 2011 IMD is adopted for the latter. I present next the OS VectorMap District dataset while I only follow with a brief recap for the OSM and IMD datasets as I have already presented them in the previous chapter.

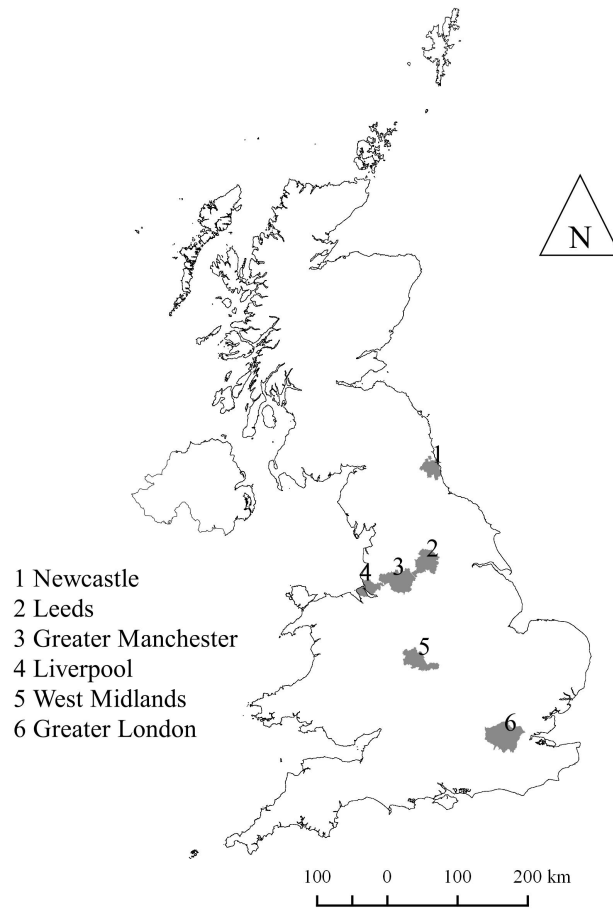


Figure 5.1: Location of the six urban areas under study.

Features of urban form are extracted from OS VectorMap District, an official digital map of the UK containing information on multiple geographic elements such as roads, building footprints, and natural resources.¹ Each road is categorised through a hierarchical system based on size (e.g., motorway, A road, B road). OS VectorMap District is produced and kept updated by Ordnance Survey, the official mapping agency of the UK, and was made openly accessible, for the first time, in April 2010. Data is provided in tiles of 100 km by 100 km. For the purpose of this analysis, I thus select the tiles pertaining to the urban areas under study. I provide a summary of the number of features (i.e., road segments, building footprints) obtained for each urban area in Table 5.2, while I present maps with this very same

¹<https://www.ordnancesurvey.co.uk/business-and-government/products/vectormap-district.html>

information for Newcastle, Leeds, and Greater Manchester in Figure 5.2 and for Liverpool, West Midlands, and Greater London in Figure 5.3. Note that the data used in this analysis dates back to September 2015.

OSM is a crowd-sourced project aimed to build an openly accessible and editable map of the world. Its accuracy has been reported to be high in different countries, especially in urban contexts [79, 80, 81]. Data for the urban areas under study was obtained through Geofabrik,² in December 2015.

IMD is an official indicator of the socio-economic deprivation of communities. It is computed through household surveys every four to five years for small census areas, called Lower-layer Super Output Areas (LSOAs). IMD is calculated by considering seven different domains: income, employment, education, health, crime, barriers to housing and services, and living environment. The deprivation index used in this analysis dates back to 2011. Although the datasets presented are not temporally aligned, I assume that the results would still be reliable as urban form changes at a very low pace.

5.3 Approach

The analytical approach is derived from the methodology presented in Chapter 3 and consists of the following steps: (i) defining metrics of the configuration of the urban environment; (ii) selecting a spatial unit of analysis and computing the metrics for such unit; (iii) normalising and scaling the metrics; (iv) testing for collinearity, to avoid overinflated regression coefficients; (v) performing regression analysis with control for spatial autocorrelation. I follow with more details for each step next.

²<http://www.geofabrik.de/index.html>

Table 5.2: Number of road segments and building footprints for the six urban areas under study.

<i>Urban area</i>	<i>Road segments</i>	<i>Building footprints</i>
Newcastle	56,456	32,193
Leeds	77,960	53,963
Greater Manchester	120,133	79,535
Liverpool	51,963	32,014
West Midlands	81,693	58,092
Greater London	165,921	102,187



Figure 5.2: Building footprints and road segments in central areas of Newcastle, Leeds, and Greater Manchester.

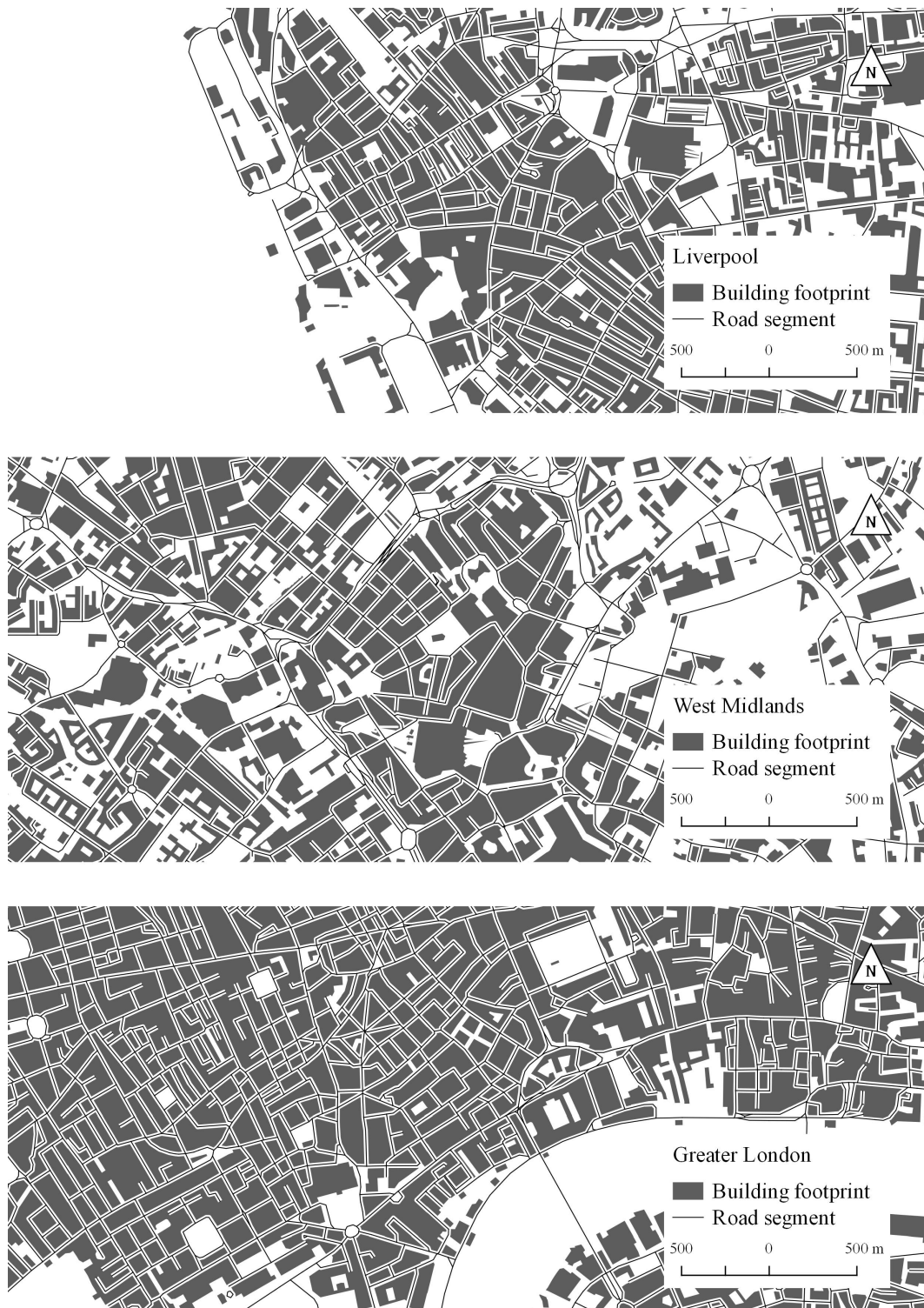


Figure 5.3: Building footprints and road segments in central areas of Liverpool, West Midlands, and Greater London.

5.3.1 Metrics

The metrics used in this analysis quantify configurational aspects of the urban environment. I define a total of nine metrics, eight are extracted from OS VectorMap District, one is extracted from OSM. Next, I present the proposed metrics, together with the theoretical justifications for their use, and a summary table (Table 5.3) with brief descriptions and formulae. The eight metrics extracted from OS VectorMap District are:

- **Connected Node Ratio** (*cnr*). It measures the level of connectivity and walkability of a street network. It is computed as the ratio between the amount of non-dead end intersections and the total amount of intersections per areal unit [101]. I include this metric as connectivity and walk-ability were considered important aspects for thriving neighbourhoods, especially from those supporting the compact city form, for example Jacobs and Gehl. In their views, these aspects not only positively affected the well-being of individuals but also enhanced social interactions, commercial activities, and provided more informal surveillance against street crimes [3, 32, 4].
- **Intersection Density** (*id*). It quantifies the density of street intersections in city areas. It is calculated as the ratio between the amount of intersections and the extension of the areal unit (in square meters) [101]. The justification for the use of this metric is similar to the one above. Intersection Density and Connected Node Ratio are, in fact, closely related as a denser street network is usually associated with more connectivity and walk-ability and thus with the positive aspects outlined above (i.e., more well-being, social and economic benefits, more safety against crime [3, 32, 4]).
- **Percentage of Unbuilt Land** (*pul*). It measures the amount of land which is left unbuilt in a city area. It is computed by dividing the amount of land without buildings by the total extension of the areal unit (in square meters), and by then multiplying this quantity by 100. Percentage of Unbuilt Land provides information on whether an area is sparsely built (i.e., high percentage of un-

built land) or densely built (i.e., low percentage of unbuilt land). I include this metric as this aspect was considered relevant by different school of thoughts. On one side, modernist planners favoured a sparser urban configuration that of the “towers in the park” [5, 6, 33]. On the other, other authors were in favour of the compact city and thus supported a denser urban form [3, 32, 4].

- **Population Density** (*pd*). It quantifies how densely populated are regions or city areas. Population Density is widely used by governments and administrations as a general statistical datum. It is calculated as the ratio between the number of people living in a specific area and the extension of such area (usually in hectares). This metric together with Percentage of Open Space provide information on how density is distributed across space. For example, if a neighbourhood has a conspicuous amount of unbuilt land and high population density, it is likely that it is characterized by residential towers. I include this metric as population density was considered a crucial aspect for city liveability from different authors, for example, Whyte [32] and Gehl [4].
- **Betweenness Centrality** (*bc*). It is based on the concept that a street segment is central if it is included in many of the shortest paths linking couple of nodes (street intersections) in a street network. The Betweenness value of the street segment $\alpha = 1, \dots, K$ is computed as follows:

$$C_{\alpha}^B = \frac{1}{(N-1)(N-2)} \sum_{j,k=1,\dots,N; j \neq k} \frac{n_{jk}(\alpha)}{n_{jk}}$$

where n_{jk} represents the number of shortest paths between nodes j and k , and $n_{jk}(\alpha)$ is the number of shortest paths between nodes j and k which contain segment α [102]. I included Betweenness Centrality as previous works found it to be associated with positive aspects of cities, such as employment density [103], concentration of retail and services [104], and street quality [105]. Betweenness Centrality is usually computed for street segments. However, since this analysis is carried out at areal level, aggregation is necessary. Betweenness Centrality is thus firstly computed at the level of street segments

and, secondly, is aggregated for areas by considering the maximum values within such areas. One can wonder whether, by doing this, I am losing spatial information or overly generalise values of Betweenness Centrality. I argue that the aggregated value, computed as the maximum value of Betweenness Centrality for an area, can be representative of the degree of accessibility of such area within the whole city system.

- **Irregularity of the Street Network (*isn*).** It quantifies the extent to which an area is characterized by a more uniform spatial configuration (e.g., grid) or by a more irregular one, with an higher variation in node degrees (i.e., number of street segments connected to an intersection). Intuitively, if an area has a more uniformly configured street network, its nodes would tend to have the same or very similar degree. This is the case, for example, of grid layouts where all the nodes have degree four, as they always have four streets converging in them. Conversely, if an area has a more irregular street network, the degrees of its nodes would tend to vary more, for example, by having cul-de-sac, three way intersections, four way intersections, six way intersections. Irregularity of the Street Network is computed by dividing the standard deviation of the node degrees of an area by the average node degree relative to such area. I include this metric as different authors considered the regularity – or irregularity – of the street network an important aspect for city liveability. For example, Jacobs generally favoured a street layout based on the grid, however, she also argued that this should be interrupted, at times, by diagonal axes and squares [3].
- **Density of Dead-end Intersections (*ddi*).** It measures to what extent a city area is characterised by the presence of dead-end roads (cul-de-sac). Density of Dead-end Intersections is calculated by dividing the number of street intersections with degree one (i.e., cul-de-sac) in an area, by the extension of such area (in square meters). This specific spatial configuration, that of the cul-de-sac, is another aspect of urban form which gained the attention of some authors. On the one hand, Jane Jacobs argued that cul-de-sac were detri-

mental to urban liveability, and in particular to safety against crimes, as they decreased street connectivity and thus the passage of pedestrians who could guarantee an informal control against these events [3]. On the other, Oscar Newman supported this spatial configuration as, in his view, the passage of strangers was associated with more crimes, thus a reduced connectivity was deemed beneficial [41].

- **Offering Advantage of Historic Properties** (*oahp*). It quantifies to what extent a city area is characterized by historic properties, compared to the average offering of such properties across the whole city. In this analysis, a property is considered historic if it is built before 1900. Offering Advantage of Historic Properties is computed through the *Offering Advantage* (OA) formula presented in the previous chapter. For the specific case of this metric, the OA formula is adapted to reflect to what extent a neighbourhood n_k offers more historic properties h_i , compared to the average city area. To be more specific:

$$OA(h_i, n_k) = \frac{\text{count}(h_i, n_k)}{\sum_{j=1}^N \text{count}(h_j, n_k)} \cdot \frac{\sum_{j=1}^N \text{count}(h_j)}{\text{count}(h_i)}$$

where $OA(h_i, n_k)$ denotes the OA of historic properties h_i in the area n_k ; $\text{count}(h_i, n_k)$ represents the total amount of occurrences of historic properties h_i in the area n_k ; N is the total number of historic properties; finally, $\text{count}(h_i)$ is the total amount of occurrences of historic properties h_i in the whole city.

Offering Advantage of Historic Properties can be considered a proxy for the traditional urban form. The more a neighbourhood offers historic properties, the more likely is that such neighbourhood has the features of the traditional compact city form (e.g., density, connectivity, perimeter blocks). Different authors had opposite perspectives on this aspect. For example, Jane Jacobs was a strenuous supporter of the traditional city form, as, in her view, it enhanced the liveability of neighbourhoods, in terms of social ties, commercial activities, and safety [3]. Modernist architects, such as Le Corbusier, on the other hand, despised the traditional city form as they saw it as overly

dense, unhealthy, and not apt to host the transport mode of the future, the car [5].

One metric is extracted from OSM, instead:

- **Percentage of Green Areas (*pga*)**. It measures the amount of greenery relative to the extension of a city area. It is computed by dividing the amount of green areas (in square meters), present in a city area, by the total extension of such area (in square meters), and by multiplying this quantity by 100. The selection of this metric is due to the relevance attributed to greenery by different authors. Jacobs, for example, argued that urban parks and gardens generally had positive effects on city liveability. However, they could also have negative impacts, especially in terms of safety, if these were not integrated in the urban fabric. For example, if they were relegated to peripheral areas with low built density [3].

Table 5.3: Metrics of the configuration of the urban environment with relative descriptions and formulae.

<i>Name of metric</i>	<i>Description</i>	<i>Formula</i>
Connected Node Ratio (<i>cnr</i>)	Level of connectivity of the street network	n^o of non dead-end intersections / total n^o of intersections
Intersection Density (<i>id</i>)	Density of street intersections	total n^o of intersections / area (m^2)
Percentage of Unbuilt Land (<i>pul</i>)	Amount of land left unbuilt	m^2 of unbuilt land / area (m^2) 100
Population Density (<i>pd</i>)	Density of city dwellers	n^o of residents / area (ha)
Betweenness Centrality (<i>bc</i>)	Level of accessibility of streets	see paragraph <i>Betweenness Centrality in Section 5.3.1</i>
Percentage of Green Areas (<i>pga</i>)	Amount of green areas	m^2 of green areas / area (m^2)
Irregularity of the Street Network (<i>isn</i>)	Level of irregularity of the street layout	SD node degree / AVG node degree
Density of Dead-end Intersections (<i>ddi</i>)	Density of dead-end roads (cul-de-sac)	n^o of dead-end intersections / area (m^2)
Offering Advantage of Historic Properties (<i>oahp</i>)	Weighted offering of historic properties	see paragraph <i>Offering Advantage of Historic Properties in section 5.3.1</i>

5.3.2 Spatial Unit of Analysis

The indications given in the general methodology suggest that the spatial unit of analysis should not be too small to render the metrics of urban form inaccurate and that, at the same time, it should keep the morphological unity of neighbourhoods, for example by not cutting blocks. For these reasons, I choose, as I did in the analysis in Chapter 4, the *ward* as spatial unit. Wards are, in fact, never too small to cause issues in the computation of the metrics of urban form and usually respect the morphological unity of city areas. As I explained in Chapter 4, wards are long standing UK administrative boundaries that represent electoral districts as well as ceremonial entities. I identify 625 wards for Greater London, 215 for Greater Manchester, 163 for West Midlands, 119 for Newcastle, 100 for Liverpool, and 93 for Leeds.

I then use the same aggregation procedure presented in Chapter 4. This consists in aggregating IMD values at ward level by averaging the IMD values associated with the LSOAs lying within each ward. As I explained before, this does not lead to a relevant loss of data as the LSOAs perfectly fit within ward boundaries and the standard deviation of the IMD values associated with them, within each ward, is very low and always smaller than their average values.

IMD and the metrics of urban form are thus computed at the level of wards. I present the summary statistics for such metrics in Table 5.4. I do not compute the False Discovery Rates (FDRs) associated with the selected metrics, in this investigation, as these are all correlated with deprivation and are very few, compared to the total number of observations for each urban area. I thus assume that the issue of multiple testing does not constitute a threat for this specific study.

5.3.3 Regression Analysis

Two steps are needed before performing a regression analysis between the proposed metrics and socio-economic deprivation. First, to meet the assumption of normality required by linear regression models, it is necessary to normalise the metrics. This can be achieved, for example, through exponentiation or logarithmic transformations. Since the proposed variables have different magnitudes (some are percentages, some others are measures of density), the regression coefficients asso-

Table 5.4: Mean and standard deviation values for the metrics of the configuration of the urban environment and IMD. NC: Newcastle; LE: Leeds; GM: Greater Manchester; LI: Liverpool; WM: West Midlands; GL: Greater London.

<i>Variable name</i>	<i>NC</i>		<i>LE</i>		<i>GM</i>	
	Mean	SD	Mean	SD	Mean	SD
Connected Node Ratio (<i>cnr</i>)	0.72	0.08	0.70	0.04	0.70	0.06
Intersection Density (<i>id</i>)	0.70	0.31	0.57	0.34	0.61	0.34
Percentage of Unbuilt Land (<i>pul</i>)	76.85	9.68	80.63	8.56	78.91	9.20
Population Density (<i>pd</i>)	43.20	17.98	40.95	22.49	41.31	17.72
Betweenness Centrality (<i>bc</i>)	0.04	0.04	0.04	0.02	0.03	0.05
Percentage of Green Areas (<i>pga</i>)	4.07	4.58	7.54	4.47	6.71	7.07
Irregularity of the Street Network (<i>isn</i>)	0.38	0.05	0.40	0.03	0.40	0.03
Density of Dead-end Intersections (<i>ddi</i>)	0.26	0.12	0.23	0.12	0.25	0.11
Offering Adv. of Historic Properties (<i>oahp</i>)	0.67	0.79	1.17	0.71	0.92	0.63
Index of Multiple Deprivation (<i>IMD</i>)	26.30	12.65	28.31	13.11	27.06	14.09

<i>Variable name</i>	<i>LI</i>		<i>WM</i>		<i>GL</i>	
	Mean	SD	Mean	SD	Mean	SD
Connected Node Ratio (<i>cnr</i>)	0.73	0.07	0.71	0.06	0.77	0.08
Intersection Density (<i>id</i>)	0.67	0.32	0.52	0.22	0.63	0.34
Percentage of Unbuilt Land (<i>pul</i>)	75.51	9.31	75.04	9.34	69.66	13.35
Population Density (<i>pd</i>)	47.72	18.48	42.47	17.11	80.09	50.35
Betweenness Centrality (<i>bc</i>)	0.06	0.11	0.03	0.02	0.02	0.01
Percentage of Green Areas (<i>pga</i>)	7.40	6.95	7.67	4.93	8.98	9.56
Irregularity of the Street Network (<i>isn</i>)	0.38	0.04	0.39	0.04	0.34	0.06
Density of Dead-end Intersections (<i>ddi</i>)	0.24	0.14	0.21	0.11	0.17	0.09
Offering Adv. of Historic Properties (<i>oahp</i>)	0.88	1.05	0.41	0.55	1.02	1.15
Index of Multiple Deprivation (<i>IMD</i>)	35.33	17.17	26.71	14.07	20.73	10.60

ciated with such values would be hard to compare and interpret. A second step thus requires scaling the normalised metrics. This can be achieved through the computation of *z*-scores.

As I explained in the previous chapter, collinearity (i.e., two or more candidate variables for a regression analysis are strongly correlated) can pose a potential threat to the reliability of a regression model. The overall fit and regression coefficients can, in fact, show over-inflated values or unexpected signs. Since this analysis is based on the interpretation of such values, I use the Variance Inflation Factor (VIF) technique to identify collinearity and discard strongly cross-correlated variables. Formula and functioning of such technique have already been presented in Chapter 4. I thus refer the reader to this section for more details. The outcome of this process is a set of variables devoid of collinearity that can be regressed against IMD, through a linear model.

A further test is needed to ascertain the statistical robustness of the findings. Such test is called Moran's test [89] and is used to check whether spatial autocorrelation is present in the residuals of a regression analysis. This phenomenon, as I explained before, can produce over-inflated regression coefficients or unexpected signs, as an effect of not considering the potential relation of observations located close to each other in space in the equation. If the outputs of this test are close to zero or they are not statistically significant, the results of the regression analysis can be accepted and then interpreted. Conversely, if the Moran's test is found to be significant, one should use a model that accounts for spatial autocorrelation. In this investigation, as in the one carried out in the previous chapter, I use the Spatial Autoregressive (SAR) model, a technique that incorporates the information associated with the location of observations through a spatial weighting matrix. The outputs of such regression should be devoid of spatial autocorrelation. However, this might not always be necessarily the case. To ascertain this aspect in a definitive way, I thus perform a second Moran's test, this time on the residuals of the SAR model. If the outcome of this test is not significant, the results of the SAR model can be trusted and then interpreted. Conversely, if the output of the test is statistically significant, the SAR model should be rejected and alternative ways of incorporating the spatial information should be searched.

5.4 Results

In this section, I first present some preliminary results drawn from the observation of the density distributions of the ten metrics considered: nine of the configuration of the urban environment and one of deprivation (as shown in Figure 5.4).³ Secondly, I illustrate the outcomes of the VIF test before presenting the results of the regression analyses. Lastly, I offer interpretations for the values and signs of the regression coefficients.

³Maps for each metric, for the six urban areas under study, are presented in Appendix B.

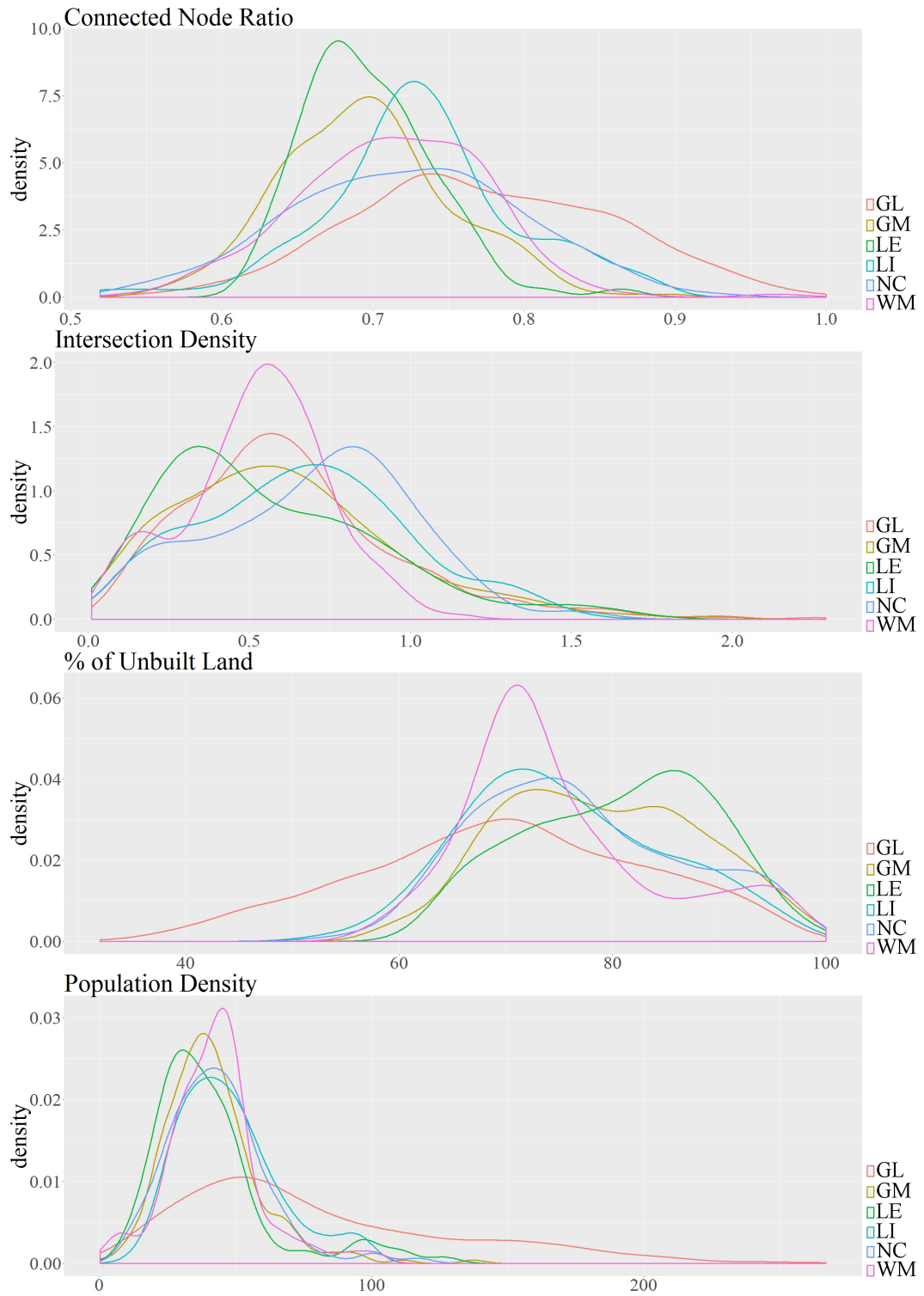


Figure 5.4: Density distributions of the metrics of urban form and IMD, for the six cities. Red corresponds to Greater London (GL); ochre to Greater Manchester (GM); green to Leeds (LE); aqua green to Liverpool (LI); blue to Newcastle (NC); purple to West Midlands (WM).

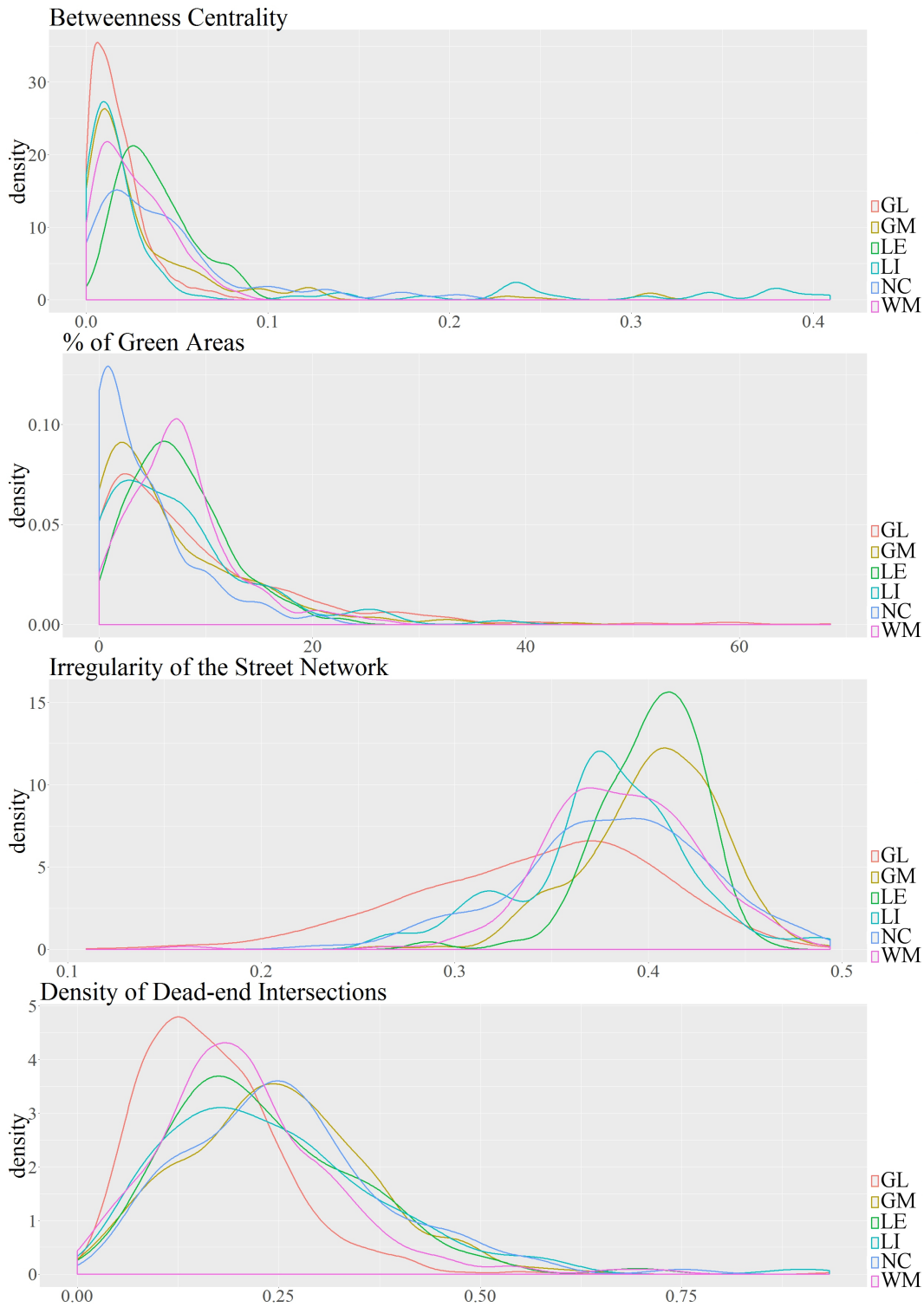


Figure 5.4: Density distributions of the metrics of urban form and IMD, for the six cities (cont.). Red corresponds to Greater London (GL); ochre to Greater Manchester (GM); green to Leeds (LE); aqua green to Liverpool (LI); blue to Newcastle (NC); purple to West Midlands (WM).

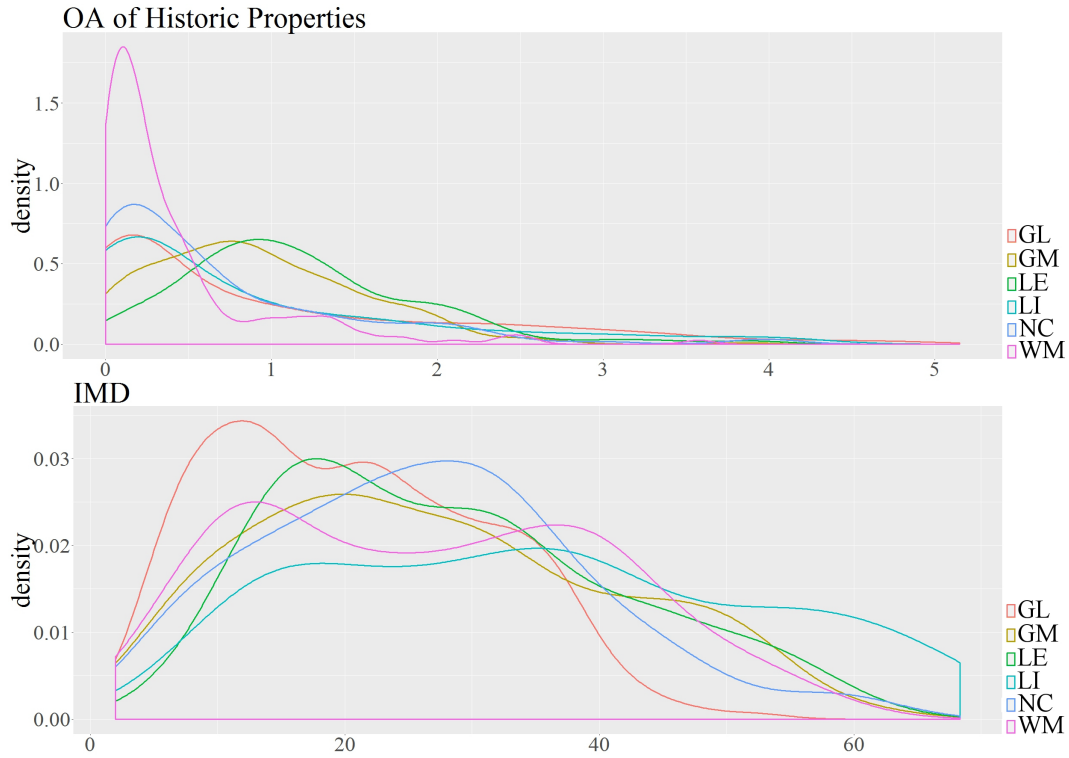


Figure 5.4: Density distributions of the metrics of urban form and IMD, for the six cities (cont.). Red corresponds to Greater London (GL); ochre to Greater Manchester (GM); green to Leeds (LE); aqua green to Liverpool (LI); blue to Newcastle (NC); purple to West Midlands (WM).

5.4.1 Preliminary results

The selected metrics do not generally show normal distributions. The only variable with a distribution close to the normal is Connected Node Ratio (*cnr*). Six metrics out of nine show a positive skew, with the majority of the values concentrating around the first quartile. These are: Percentage of Green Areas (*pga*), Intersection Density (*id*), Density of Dead-end Intersections (*ddi*), Population Density (*pd*), Betweenness Centrality (*bc*), and Offering Advantage of Historic Properties (*oahp*). Two variables out of nine present a negative skew instead, with most of the values clustering around the third quartile. These are: Irregularity of the Street Network (*isn*) and Percentage of Unbuilt Land (*pul*). IMD also shows a non-normal distribution as tends to show a positive skew. Graphs of the density distributions of each metric, for each of the six cities under study, are presented in Figure 5.4.

Some remarks can be drawn from these preliminary results. First, London's

urban form (colour coded red in Figure 5.4) seems to differ from that of the other cities. It tends to be denser in terms of built form and population (i.e., more low values of Percentage of Unbuilt Land, more high values of Population Density compared to the other cities). Moreover, London also appears to have a better connected street network: more high values of Connected Node Ratio, more low values of Density of Dead-end Intersections compared to the other cities. Second, West Midlands' urban features (colour coded purple in Figure 5.4) seems to deviate quite substantially from the other cities. It tends to show peaks of values rather than more varied distributions. Most of its wards tend to have low values of Intersection Density (around 0.5), moderately high values of Percentage of Unbuilt Land (around 70%), and a poor offering of historic properties, with most of the values of the metric Offering Advantage of Historic Properties close to 0. Another exception to the general trend is Newcastle (colour coded blue in Figure 5.4), with a relative low offering of green areas. Most of its wards tend to have, in fact, values of the metric Percentage of Green Areas close to zero. Finally, Leeds (colour coded green in Figure 5.4) seems to be more sparsely built (i.e., more high values of Percentage of Unbuilt Land) and offer more historic properties (i.e., more high values of Offering Advantage of Historic Properties) compared to the other cities. Note that this last aspect seems to be valid also for Greater Manchester (colour coded ochre in Figure 5.4). For what concerns deprivation, Greater London (colour coded red in Figure 5.4) appears to be the most advantaged urban area of the set, with most of the IMD values concentrating around 10 and taking values above 40 in few cases only. Conversely, Liverpool (colour coded aqua green in Figure 5.4) seems to be the least advantaged, with a long tail of values above 50.

5.4.2 Modelling Urban Form and Deprivation Across the Six Cities

Having performed the normalisation and scaling procedures illustrated in the previous section, I use the Variance Inflation Factor (VIF) technique to detect and discard variables that show strong collinearity. Outcomes of such method do find that some of the candidate variables show the issue. In particular, strong collinearity is de-

tected, in all case studies, for Connected Node Ratio (*cnr*) and Intersection Density (*id*), with VIFs significantly greater than 10. Similarly, a VIF value higher than 10 is output for Percentage of Unbuilt Land (*pul*), in Leeds. The VIF technique thus discards such variables from the list of candidates for the regression analysis.

I thus input the remaining variables in six regression models, one for each of the cities considered, with IMD as dependent variable. Results (see Table 5.5) suggest that the selected features of urban form are associated with levels of socio-economic deprivation of UK city dwellers. The models are statistically significant (at 99% confidence level) and generally present moderate fits, with four cities (i.e., West Midlands, Greater London, Greater Manchester, Leeds) out of six showing adjusted R^2 values around 0.50. To be more specific, urban form can explain 50% of the variance of IMD in West Midlands, 49% in Greater London, 48% in Greater Manchester, and 50% in Leeds. The adjusted R^2 values for Liverpool and Newcastle are lower, instead. Urban form can explain 25% of the variance of IMD in Liverpool and only 11% in Newcastle. To check whether residuals do not show spatial autocorrelation, I perform the Moran's test. Outputs show that there is statistical evidence of the presence of spatial autocorrelation in all models. Moran's I values are statistically significant (at 99% confidence level) and vary between a minimum of 0.16 (Greater Manchester, Leeds, and Newcastle), to a maximum of 0.44 (Greater London). I thus use the Spatial Autoregressive (SAR) technique to account for the spatial information associated with the observations. SAR models are all statistically significant (at 99% confidence level) and show greater adjusted R^2 values and smaller coefficients, meaning that part of IMD is indeed explained by the spatial factor. More specifically, the SAR model for West Midlands can explain 67% of the variance of deprivation, the one for Greater London can explain 70%, the one for Greater Manchester 56%, the one for Leeds 59%, the one for Liverpool 49%, while the one for Newcastle can explain 27% of the variance of IMD. A second Moran's test, performed on the residuals of the SAR models, highlights that there is no statistical evidence of the presence of spatial autocorrelation (i.e., p -value > 0.05 , in all cases). Results of such regressions are thus reliable. As for the regression coef-

ficients, I observe common patterns of signs, strengths, and significance across the six cities:


















- Density of Dead-end Intersections (*ddi*) is significant and positively associated with deprivation, in five cities out of six (i.e., West Midlands, Greater London, Greater Manchester, Liverpool, and Newcastle);
- Irregularity of the Street Network (*isn*) is significant and negatively associated with deprivation, in four cities out of six (i.e., West Midlands, Greater Manchester, Liverpool, and Newcastle);
- Percentage of Unbuilt Land (*pul*) is significant and positively associated with deprivation, in four cities out of six (i.e., Greater London, Greater Manchester, Liverpool, and Newcastle);
- Population Density (*pd*) is significant and positively associated with deprivation, in four cities out of six (i.e., West Midlands, Greater London, Greater Manchester, and Leeds).


















For what concerns the remaining coefficients, Betweenness Centrality is associated with more deprivation in two cities out of six (i.e., Greater London and Leeds), Offering Advantage of Historic Properties is related to less deprivation in Greater London only, while Percentage of Green Areas is negatively associated with deprivation in Newcastle only. I present the full results of the linear regression (LR) and SAR models in Table 5.5.






















5.4.3 Interpretations

As I illustrated in the previous section, several regression coefficients (i.e., Density of Dead-end Intersections, Irregularity of the Street Network, Percentage of Unbuilt Land, and Population Density) show similar patterns across the six cities under study, meaning that these aspects of urban form are associated with socio-economic deprivation at country level. I thus argue that neighbourhoods with high levels of socio-economic deprivation of urban England are characterised by high population density, vast unbuilt surface, strong presence of dead-end roads, and regular street

Table 5.5: LR and SAR models for the six urban areas under study. Red bar means positively associated with IMD, blue bar means negatively associated with IMD. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘**’ significant at $p < 0.01$; ‘***’ significant at $p < 0.001$. The configuration of the urban environment can explain up to 70% of the variance of IMD. Density of Dead-end Intersections (*ddi*), Irregularity of the Street Network (*isn*), Percentage of Unbuilt Land (*pul*), and Population Density (*pd*) appear to be the most important drivers of deprivation, across all cities.

Ind. var.	Greater London				Greater Manchester			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.01		0.00		-0.02
<i>cnr</i>	-		-		-		-	
<i>id</i>	-		-		-		-	
<i>pul</i>	***	0.36 	***	0.27 	***	0.91 	***	0.78 
<i>pd</i>	***	1.06 	***	0.60 	***	0.35 	***	0.27 
<i>bc</i>	**	0.10 	.	0.04		-0.05		-0.07
<i>pga</i>		0.01		-0.01		0.06		0.04
<i>isn</i>	*	0.11 		0.03	***	-0.53 	***	-0.46 
<i>dd</i>		0.02	*	0.08 	***	0.96 	***	0.85 
<i>oahp</i>	***	-0.17 	***	-0.14 		-0.06		-0.05
adj. R^2		0.49		0.70		0.48		0.56
p-val.		0.00		0.00		0.00		0.00
Moran's		0.04		0.03		0.16		0.02
p-val.		0.17		0.08		0.00		0.26

Ind. var.	West Midlands				Leeds			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.05		0.00		-0.01
<i>cnr</i>	-		-		-		-	
<i>id</i>	-		-		-		-	
<i>pul</i>		0.12 		0.13 	-		-	
<i>pd</i>	***	0.45 	***	0.35 	**	0.44 	**	0.37 
<i>bc</i>	**	0.19 		0.06	**	0.21 	*	0.17 
<i>pga</i>	*	0.14 		0.07		-0.10		-0.07
<i>isn</i>	*	-0.23 	*	-0.17 		-0.10		-0.10
<i>dd</i>	*	0.24 	*	0.18 	.	0.23 		0.19 
<i>oahp</i>	.	0.11 		0.06		-0.04		-0.05
adj. R^2		0.50		0.67		0.50		0.59
p-val.		0.00		0.00		0.00		0.00
Moran's		0.30		0.03		0.16		-0.02
p-val.		0.00		0.21		0.00		0.59

Ind. var.	Liverpool				Newcastle			
	LR		SAR		LR		SAR	
	p-val.	β	p-val.	β	p-val.	β	p-val.	β
(intercept)		0.00		-0.02		0.00		0.01
<i>cnr</i>	-		-		-		-	
<i>id</i>	-		-		-		-	
<i>pul</i>	.	0.40 	*	0.36 	*	0.60 	*	0.50 
<i>pd</i>	.	0.33 		0.18 		0.10		0.07
<i>bc</i>		0.14 		0.07	.	0.16 		0.13 
<i>pga</i>		-0.06		-0.02	*	-0.20 	.	-0.14 
<i>isn</i>	.	-0.28 	.	-0.26 	.	-0.38 	.	-0.34 
<i>dd</i>	**	0.66 	**	0.58 	**	0.62 	**	0.53 
<i>oahp</i>		-0.11 		-0.12 		0.07		0.05
adj. R^2		0.25		0.49		0.11		0.27
p-val.		0.00		0.00		0.01		0.00
Moran's		0.25		0.01		0.16		-0.02
p-val.		0.00		0.37		0.00		0.56

patterns. Such urban configuration closely resembles the modernist “towers in the park” approach, which consisted of tower blocks (many residents concentrated in a small portion of land) laid out in open space, reached by dead-end roads, and surrounded by a regular and repetitive street pattern (see, for example, [5, 6, 33]).

The link between this urban configuration and deprivation seem to corroborate the theories of the compact city form (see, for example, [3, 32, 4]). Jacobs, for example, supported perimeter blocks rather than tower blocks laid out in open space as the retraction of buildings from the line of side-walks diminished social interactions, as fewer points of exchange between streets and buildings (e.g., doors, windows, porches, porticoes) were present. Furthermore, they reduced commercial activities, as there was no physical space along streets where to place them, and safety, as streets were missing the informal supervision granted by windows facing them (the so-called “eyes on the street” effect) [3]. For similar reasons, she also favoured well-connected street networks rather than dead-end roads, as the latter reduced connectivity and thus the ability of pedestrians to navigate the urban space. She deemed this an important aspect not only for the social vitality of streets but also for their economic prosperity and safety. Fewer passers-by corresponded, in her view, to fewer chances of social interactions, a smaller likelihood of people shopping, and less informal control against crime [3]. Finally, she favoured street networks with some irregularities rather than overly regular ones. In her view, the latter had the negative effect of jeopardising “visual interruptions” (e.g., diagonal roads, squares), an aspect that enhanced urban life [3].

The findings of this analysis seem thus to disagree with the modernist theories on urban form (see, for example, [5, 6, 33]). Similarly, the theory advanced by Newman (i.e., dead-end roads were beneficial against crime as they reduced passage of people and thus created more controllable spaces [41]) seems to be invalidated. Although I did not specifically test a measure of crime, a domain that quantifies such issue is included in the IMD score.

As for the remaining significant relationships, the link between Betweenness Centrality and deprivation can be due to the scale at which it is computed (i.e., en-

tire cities). Previous studies have shown that this measure correlated with positive aspects such as presence of commercial activities and services [104], employment density [103], and street quality [105]. However, Xiao found that this very same metric was also associated with roads with high traffic volume, when computed at the scale of an entire city [106]. I suggest that this could be the case for this analysis too. Moreover, I argue that roads with high traffic volume are also associated with higher levels of noise and pollution. This, in turn, might be detrimental to people's living conditions and thus be associated with deprivation. The inverse relationship between Offering Advantage of Historic Properties and deprivation in Greater London seems to be in line with the theories supporting the traditional compact city form [3, 32, 4]. This metric, in fact, can be considered a proxy for the traditional urban form as the above-average presence of historic properties are likely to be associated with features such as perimeter blocks, density, and connectivity. Finally, the negative association between Percentage of Green Areas and deprivation in Newcastle is backed up by studies in the Public Health domain. For example, Maas *et al.* reported that the percentage of green areas in the urban environment had a positive relation with the perceived health of city dwellers [107].

5.5 Discussion

In this chapter, I applied the methodology presented in Chapter 3 to study the relationship between configurational aspects of the urban environment and levels of socio-economic deprivation for six UK cities (i.e., Greater London, Greater Manchester, West Midlands, Liverpool, Leeds, and Newcastle). The metrics were extracted from two openly accessible datasets, namely OS VectorMap District and OSM, while the socio-economic levels were obtained from the 2011 IMD. I used regression analysis to model the relationship between metrics of urban form and levels of socio-economic deprivation of the six cities under study and I interpreted the outcomes. Results suggest that configurational features of the urban environment could explain the socio-economic levels of the selected urban areas up to a certain extent: four models out of six (i.e., West Midlands, Greater London, Greater

Manchester, Leeds) were able to explain up to 70% of the variance of the socio-economic index. For what concerns regression coefficients, high population density, vast portions of unbuilt land, high density of dead-end roads, regularity of the street network were the most common predictors across all cases, with the first three being positively associated with deprivation, while the last being inversely related to it. The typical deprived neighbourhood of urban England seems thus to resemble the so-called “towers in the park” modernist development, a type of plan characterized by residential tower blocks detached from the side walks, a regular and repetitive street pattern, and presence of dead-end roads. The urban features captured by this analysis seem to be in line with what authors supporting the compact city form deemed detrimental to urban life.

5.5.1 Limitations

I ought to acknowledge some limitations for this study. A first limitation concerns the outcomes of the regression analyses. I found four metrics to be highly descriptive of IMD, across all cities. Although this is an improvement over previous studies that focused on one variable only, more metrics should be included to have a more detailed characterisation of what kind of urban configuration is linked to socio-economic levels of citizens and to improve the explanatory power of the models. For example, one can include information on external features of buildings (e.g., material, surface window area).

A second limitation specifically regards the models for Liverpool and Newcastle, which under performed compared to the other cities. The issue seems not to lie neither in the paucity of data nor in a small sample size, rather, it seems to be related to their size. Liverpool and Newcastle are, among the selected urban regions, the smallest in terms of population and extension. I argue that, in smaller cities, urban form may not be as relevant in explaining socio-economic levels as is in bigger ones as different factors may be at stake (e.g., education, proximity to economic centres).

A third limitation regards the generalizability of the findings. At the moment, I can only claim that results hold for the six urban areas under study and for 2011 (i.e., the year for which IMD is valid). It is not possible to claim validity of the

findings across different time spans and for different urban areas. Nonetheless, the approach presented in this chapter can be applied to other geographic contexts and periods to check whether findings hold or not.

A fourth limitation concerns causality. The approach presented in this chapter is purely based on correlation and regression analysis and thus does not establish a causal relationship between the phenomena analysed.

In this chapter and in the previous one, I applied the proposed methodology to test the two hypotheses of this thesis, taken separately: amenities are associated with aspects of liveability and configurational aspect of the urban environment are related to aspects of liveability. In the next chapter, I apply the very same methodology to check whether both hypotheses, taken together, hold for Greater London and three different liveability indexes, namely deprivation, life expectancy, and childhood obesity.

Chapter 6

On the Relationship between Amenities, Configuration of the Urban Environment and Three Liveability Indexes

In Chapter 4, I used the proposed methodology to test the first hypothesis of this thesis (i.e., amenities are related to aspects of liveability) on deprivation levels of three UK cities. Outcomes showed that specific categories of amenities were associated with deprivation. Furthermore, to improve the readability of the outcomes, I grouped such categories in broader topics through thematic analysis. The topic *sport & recreation* was found, for example, to be associated with more advantaged neighbourhoods, while *unhealthy food* with more deprived ones.

In Chapter 5, I utilised the proposed methodology to test the second hypothesis of this thesis (i.e., the configuration of urban form is associated with aspects of liveability) on deprivation levels of six UK cities. Results show that specific aspects of the configuration of urban form related to deprivation levels of such cities, with the most common aspects being population density, vast portions of unbuilt land, regularity of the street network, and density of dead-end roads.

In this chapter, I use the proposed methodology to test both hypotheses together, on deprivation levels and on two other aspects of liveability, namely life ex-

pectancy and childhood obesity, for the metropolitan region of London (i.e., Greater London). This chapter is structured as follows. I first provide a brief introduction, as the core argument of this analysis is basically the sum of the ones presented in Chapter 4 and 5. I follow by illustrating the datasets from which to extract the various metrics, before illustrating how I adapt the methodology proposed in this thesis to fit the purpose of this investigation. Finally, I present outcomes and interpretations.

6.1 Introduction

Many different works tried to identify what characteristics of urban form fostered the well-being of city dwellers. Some authors favoured the traditional compact city form [3, 32, 4]. Some others, instead, supported a new way of planning based on tower blocks and a more car-oriented urban environment [5, 6, 33]. The main limitation of both these sets of works is that they were carried out through a qualitative approach and thus methodologies are hardly replicable and generalizable. More recently, advances in technology, made possible to study urban form from a quantitative perspective. Some researchers focused on specific aspects of the configuration of streets. For example, Timothee *et al.* found that more accessibility was associated with more economic activities [104]. Remali *et al.* reported that the same metric was also associated with street quality [105]. Other researchers in the field of Health focused on specific amenities, instead. Giles-Corti and Donovan, for example, found that the presence of golf courses was associated with more advantaged neighbourhoods in Australia [16]. Cummins *et al.* reported a strong relationship between the presence of fast food restaurants in England and Wales and levels of deprivation of neighbourhoods [17]. Although the methods used in these works were replicable and their outcomes more generalizable, the focus was on single aspects of urban form. However, it might well be that the relationship between urban form and any liveability index would be better described if more aspects of the urban environment are considered, as the Urban Morphology discipline proposes.

In this analysis, I use the methodology presented in Chapter 3 to quantitatively analyse the relationship between amenities, configurational aspects of the urban environment and several aspects of city liveability, namely deprivation, life expectancy, and childhood obesity, for the metropolitan area of London. Carrying out this analysis requires a four-step process. First, I define the metrics capturing the two aspects of urban form under study (i.e., amenities, configuration of the urban environment) and extract them from openly accessible datasets. Second, I normalise and scale relevant metrics. Third, I perform regression analysis with the scaled metrics as independent variables and the liveability indexes as dependent ones. Finally, I interpret the outcomes of such regressions. I follow next with a presentation of the datasets used.

6.2 Datasets

To carry out this analysis, three different datasets are necessary: one that provides information on what kind of amenities are present in London, one that contains vectorial representations of roads and buildings, and one that provides the required indexes of liveability for the area under study (i.e., Greater London). I use Foursquare and OpenStreetMap (OSM) as data sources for amenities, OSM and Ordnance Survey (OS) VectorMap District for roads and buildings, the 2011 English Index of Multiple Deprivation (IMD) for information on the socio-economic status of London citizens, and the London Ward Well-Being Scores dataset for information on life expectancy and childhood obesity of London communities. Since I have already presented all the datasets except for the London Ward Well-Being Scores, I only explain in detail the latter.

The London Ward Well-Being Scores¹ represent a combined index of well-being indicators of the resident population based on twelve different domains. These span nine different themes: health, economic security, safety, education, children, families, public transport accessibility, environment, and happiness. The London Ward Well-Being Scores are computed for wards, UK electoral districts.

¹<https://data.london.gov.uk/dataset/london-ward-well-being-scores>

Given the strong overlap between well-being and IMD, I do not consider the main index in this analysis. Rather, I consider two of its sub-domains, life expectancy and childhood obesity, which previous research found to be associated with urban form (see for example [108, 109]). This can provide deeper insights on the relationship between urban form and liveability as the focus is on specific components of the composite score.

The life expectancy score varies between -22.3 and 40.7, with greater values being associated with longer life expectancy, while smaller ones with shorter life expectancy. The childhood obesity score varies between -33.1 and 25.5, with greater values being related to less childhood obesity, while negative ones with more childhood obesity. The dataset used in this analysis dates back to 2013. I present in Figure 6.1 and Figure 6.2 the life expectancy and childhood obesity maps of Greater London.

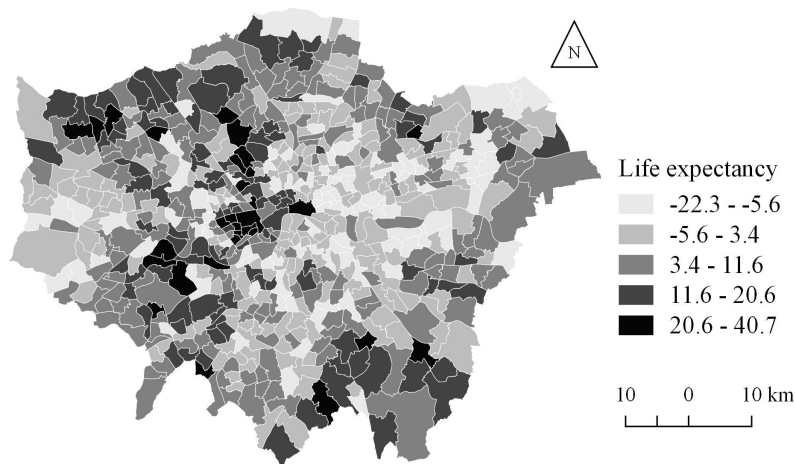


Figure 6.1: Life expectancy scores for Greater London.

Note that the datasets used in this analysis are not temporally aligned. OSM and OS VectorMap District date back to September 2015, Foursquare to April 2014, while life expectancy and childhood obesity to 2013. Nonetheless, I assume that this temporal discrepancy would not constitute a threat to the validity of the outcomes as the subject of this analysis (i.e., urban form) change at a relatively slow pace.

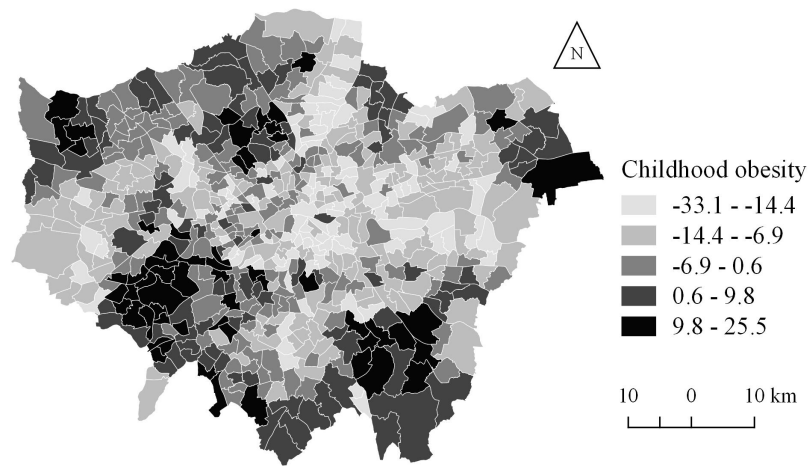


Figure 6.2: Childhood obesity scores for Greater London.

6.3 Approach

6.3.1 Metrics

I apply the *Offering Advantage* (OA) formula presented in Chapter 4 to all the Foursquare and OSM categories of amenities of London. I thus obtained 337 OA values of Foursquare categories of amenities, and 169 OA values of OSM categories of amenities. As for the metrics of the configuration of the urban environment, I use the ones presented in Chapter 5: Connected Node Ratio (*cnr*), Intersection Density (*id*), Percentage of Unbuilt Land (*pul*), Population Density (*pd*), Betweenness Centrality (*bc*), Irregularity of the Street Network (*isn*), Density of Dead-end Intersection (*ddi*), Offering Advantage of Historic Properties (*oahp*), and Percentage of Green Areas (*pga*). I present a list of all the metrics used in this analysis, with relative descriptions and formulae, in Table 6.1.

6.3.2 Spatial Unit of Analysis

The spatial units for this analysis are, once again, the *wards* (i.e., UK electoral and ceremonial boundaries). I identify 625 wards for Greater London. Since IMD is provided for spatial units (i.e., Lower-layer Super Output Areas) that are smaller than wards, I aggregate the values associated with such units at the level of wards, through the computation of their average values. Life expectancy and childhood obesity scores do not need aggregation as they are already provided for the chosen

Table 6.1: Metrics of amenities and configuration of the urban environment with relative descriptions and formulae.

<i>Name of metric</i>	<i>Description</i>	<i>Formula</i>
Offering Advantage of amenity	Weighted offering of a specific amenity	<i>see paragraph Metrics in section 4.3.1</i>
Connected Node Ratio (<i>cnr</i>)	Level of connectivity of the street network	n^o of non dead-end intersections / total n^o of intersections
Intersection Density (<i>id</i>)	Density of street intersections	total n^o of intersections / area (m^2)
Percentage of Unbuilt Land (<i>pul</i>)	Amount of land left unbuilt	m^2 of unbuilt land / area (m^2) 100
Population Density (<i>pd</i>)	Density of city dwellers	n^o of residents / area (ha)
Betweenness Centrality (<i>bc</i>)	Level of accessibility of streets	<i>see paragraph Betweenness Centrality in Section 5.3.1</i>
Percentage of Green Areas (<i>pga</i>)	Amount of green areas	m^2 of green areas / area (m^2)
Irregularity of the Street Network (<i>isn</i>)	Level of irregularity of the street layout	SD node degree / AVG node degree
Density of Dead-end Intersections (<i>ddi</i>)	Density of dead-end roads (cul-de-sac)	n^o of dead-end intersections / area (m^2)
Offering Advantage of Historic Properties (<i>oahp</i>)	Weighted offering of historic properties	<i>see paragraph Offering Advantage of Historic Properties in section 5.3.1</i>

spatial unit of analysis.

6.3.3 Correlation Analysis

The OA formula is computed at the level of wards for each OSM and Foursquare category of amenity. Since these are hundreds and might not all be related to the indexes considered in this analysis, I use the method presented in Section 4.3 to identify those that are associated with such indexes. In brief, this method requires the use of the technique by Clifford *et al.* [85] to test the correlations between each OA value of category of amenity and each of the three indexes. Furthermore, since this procedure involves the testing of hundreds of values, I compute the False Discovery Rates (FDRs) [86] to detect and discard correlations that are false positives.

6.3.4 Regression Analysis

Before performing a regression analysis between the relevant OA values, metrics of the configuration of the urban environment and the liveability indexes, I normalise and scale such values. Furthermore, to avoid collinearity among candidate vari-

ables, I use the Variance Inflation Factor (VIF) technique [88] to detect and discard variables that show collinearity. At this point, since the OA values that can potentially pass the previous step are hundreds, I regress the OA values and the metrics of the configuration of urban form against the liveability indexes through a stepwise technique, a method that discards independent variables that are only marginally associated with the dependent one. The output of this process is a model with a parsimonious set of variables strongly related to the liveability indexes. Finally, I check for the presence of spatial autocorrelation in the residuals, through the Moran's test [89]. If there is no statistical evidence of the presence of this phenomenon, the outputs of the regression can be trusted and interpreted. Conversely, if spatial autocorrelation is present, I use the Spatial Autoregressive (SAR) technique to account for such phenomenon. I then perform a second Moran's test on the residuals of the SAR model to ascertain that spatial autocorrelation is definitely absent.

It is plausible that more life expectancy and less childhood obesity are associated with greater earnings. An higher income might be, in fact, a precondition for a better living environment and more health services. Rodgers, for example, performed a statistical analysis using cross-sectional data from more than 50 countries and reported that income distribution is significantly and strongly related to mortality [110]. Rodgers also found that, in countries with high levels of inequality, life expectancy could be between five and ten years lower than in more egalitarian countries. For what concerns the link between childhood obesity and income, Wang found, for example, that the majority of subjects with low socio-economic status in the US were more at risk of childhood obesity than higher income ones [111]. Studying the relationship between urban form and life expectancy or childhood obesity thus requires accounting for income. To do so, I first regress income deprivation² against such indexes, through linear regression or, if spatial autocorrelation were present, SAR technique. I then apply my methodology using the residuals of such regression as dependent variable. In other words, I investigate to what extent urban form can explain the variance of the liveability indexes that is not

²One of the sub-domains of IMD.

already explained by income.

6.4 Results

In this section, I first present the results for the regression performed with IMD as dependent variable. Second, I illustrate the outcomes for the regressions performed with life expectancy and childhood obesity as dependent variables.

6.4.1 Amenities, Configuration of Urban Form and IMD

The selection process used to identify OA values of Foursquare and OSM categories of amenities significantly correlated with IMD output twelve metrics. Some of these relate to food provision (e.g., fried chicken restaurants), some others to sport facilities (e.g., golf courses).

I normalise and scale these OA values and the nine metrics of the configuration of the urban environment. I then use the Variance Inflation Factor (VIF) technique to check the candidate variables for collinearity. Such test detected and discarded Connected Node Ratio (*cnr*), as it had a value of 11.1 (the acceptance threshold for this test is 10). I then use stepwise linear regression to obtain a parsimonious model with the smallest number of variables possible. This discarded five variables: OA of Bus Stops, OA of Lakes, OA of Golf Courses, Percentage of Unbuilt Land, and Irregularity of the Street Network. The output of the stepwise regression is a model with fifteen metrics, all statistically relevant, at 95% confidence threshold, which can explain 54% of the variance of IMD. The strongest regression coefficients are: Population Density ($\beta = 0.63$), Intersection Density ($\beta = -0.25$), and OA of Dentist's Offices ($\beta = -0.15$). To ascertain whether these results are robust against the phenomenon of spatial autocorrelation, I perform the Moran's test on the residuals of the stepwise regression. Outputs show statistical evidence of the presence of such phenomenon (i.e., Moran's $I = 0.34$, p -value = 0.00). I thus implement the SAR technique to incorporate the spatial factor in the model. The adjusted R^2 increases to 0.76, while the coefficients become smaller and slightly less statistically significant. Population Density still is the strongest coefficient; however, it shows a smaller value ($\beta = 0.32$), while the second strongest becomes Density of Dead-end

Intersections ($\beta = 0.11$) and the third strongest Intersection Density ($\beta = -0.10$). I perform the Moran's test on the residuals and outcomes show no statistical evidence of the presence of spatial autocorrelation (i.e., Moran's $I = 0.00$, p -value = 0.41). I present full results for both linear regression (LR) and SAR model in Table 6.2.

Table 6.2: LR and SAR models for IMD. Red bar means positively associated with IMD, blue bar means negatively associated with IMD. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘**’ significant at $p < 0.01$; ‘***’ significant at $p < 0.001$. The combination of Foursquare categories and urban form can explain up to 76% of the variance of IMD, with the strongest coefficients being *pd* (i.e., $\beta = 0.32$) and *ddi* (i.e., $\beta = 0.11$).

Independent variable	LR		SAR	
	<i>p</i> -value	β	<i>p</i> -value	β
(intercept)		0.00		-0.01
<i>African restaurant</i>	**	0.10	*	0.04
<i>bank</i>	***	-0.10	***	-0.08
<i>cricket</i>	**	-0.08	**	-0.05
<i>dentist</i>	***	-0.15	***	-0.09
<i>factory</i>	***	0.14	***	0.08
<i>fried chicken</i>	***	0.13	***	0.08
<i>grocery store</i>	***	0.11	**	0.06
<i>mosque</i>	**	0.09	**	0.06
<i>salon barbershop</i>	**	-0.08	*	-0.05
<i>pga</i>	*	0.06	*	0.05
<i>ddi</i>	***	0.14	***	0.11
<i>id</i>	***	-0.25	**	-0.10
<i>bc</i>	***	0.13		0.02
<i>oahp</i>	***	-0.10	*	-0.07
<i>pd</i>	**	0.63	***	0.32
Adjusted R^2		0.54		0.76
<i>p</i> -value		0.00		0.00
Moran's I		0.43		0.00
<i>p</i> -value		0.00		0.41

6.4.2 Interpretations

To interpret the coefficients (i.e., categories of amenities), I use the categorization obtained through thematic analysis and backed up by previous studies illustrated in Chapter 4. The typical deprived London neighbourhood seems thus to be characterised by an above-average offering of *unhealthy food* (*African restaurant* and *fried chicken*) and *unpleasant places* (*factory*). Conversely, the above-average offering of *sport & recreation* facilities (*cricket*) and of amenities offering *surplus goods* (*dentist*, *salon barbershop*) seem to characterise more advantaged neighbourhoods. For

what concerns the remaining categories of amenities, the inverse relationship between *bank* and IMD might be linked to the fact that more affluent communities use financial services more than less affluent ones. However, I do not have literature to back up this hypothesis. The positive relationship between *mosque* and deprivation might be linked to the fact that Muslim communities, who tend to be less advantaged, live near their places of worship. A previous study seems to support the first hypothesis. Brimicombe found, in fact, a consistent relationship between high concentration of Muslim residents and IMD values below the median [112]. To check whether the second part of my previous statement is correct (i.e., Muslim communities live near their places of worship), I perform a correlation analysis, through the method by Clifford *et al.*, to check the relationship between the percentage of Muslim residents in London wards³ and the OA value of Mosques. The outcome is congruent with the second part of the statement as there is a statistically significant correlation between the two variables (i.e., Spearman's $r_s = 0.40$, $p\text{-value} = 0.01$). For what concerns *grocery store*, I cannot provide any explanation for the link between presence above-average of such amenity and deprivation.

As for the configuration of the urban environment, it appears that London neighbourhoods characterised by greater values of Offering Advantage of Historic Properties (*oahp*) and Intersection Density (*id*) are less deprived. Both variables describe aspects of the traditional compact city form. The former is usually associated with more connectivity, walk-ability, and human scale urban environments, the latter is related to a dense urban fabric. Population Density (*pd*) is an aspect of the traditional city form too; however, it is associated with deprivation in this analysis. This might not necessarily mean that density is completely detrimental. It might mean, instead, that it is detrimental after a certain threshold. Finally, neighbourhoods with higher Density of Dead-end Intersections (*ddi*) and higher Percentage of Green Areas (*pga*) tend to be deprived. These findings seem to be in line with the works of authors who favoured the traditional compact city form. Jacobs, Gehl, and Whyte, for example, all supported human-scale neighbourhoods characterised



³This information is extracted from the UK Census 2011.

by high connectivity, walk-ability, medium densities, and presence of historic buildings [3, 32, 4]. In particular, Jacobs argued that such features had positive effects on social interactions, commercial activities, and safety against crime. In her view, a dense and walk-able urban environment promoted the everyday face to face interactions, shopping, and use of services. Furthermore, density and walk-ability also provided a sort of informal control against crime in two ways: windows facing the streets and more people on the side-walks [3]. The association between higher Percentage of Green Areas and deprivation might be due to the fact that neighbourhoods with many green areas lose density and thus also the positive aspects associated with it and illustrated above.

6.4.3 Amenities, Configuration of Urban Form, Life Expectancy and Childhood Obesity





In this section, I first present the results of the regression analysis performed for life expectancy, second, I present the results of the one performed for childhood obesity, and, third, I offer interpretations for both models.

Table 6.3: LR and SAR models for life expectancy. Red bar means negatively associated with life expectancy, blue bar means positively associated with life expectancy. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘***’ significant at $p < 0.01$; ‘****’ significant at $p < 0.001$. Income deprivation can explain almost half (i.e., 47%) of the variation of life expectancy.

<i>Independent variable</i>	<i>LR</i>		<i>SAR</i>	
	<i>p</i> -value	β	<i>p</i> -value	β
(intercept)		0.00		0.00
<i>income deprivation</i>	***	-0.66 	***	-0.54 
Adjusted R^2		0.43		0.47
<i>p</i> -value		0.00		0.00
Moran’s I		0.21		0.06
<i>p</i> -value		0.00		0.01

Outputs of the linear regression (LR) with income deprivation as independent variable and life expectancy as dependent one show that the former can explain 43% (adjusted $R^2 = 0.43$) of the variance of the latter, with an high statistical confidence (i.e., 99%) (see Table 6.3). However, the Moran’s test performed on the residuals of the linear regression shows a significant presence of spatial autocorrelation (i.e., Moran’s $I = 0.21$, p -value = 0.00). I thus use the SAR technique to include

Table 6.4: LR model for the residuals of the regression between income deprivation and life expectancy. Red bar means negatively associated with life expectancy, blue bar means positively associated with life expectancy. ‘.’ significant at $p < 0.1$; ‘*’ significant at $p < 0.05$; ‘**’ significant at $p < 0.01$; ‘***’ significant at $p < 0.001$. Urban form can only explain up to 4% of the variance of the residuals of the regression between income deprivation and life expectancy. The strongest coefficients are *cnr* (i.e., $\beta = 0.22$) and *pul* (i.e., $\beta = 0.16$).

Independent variable	LR	
	p-value	β
(intercept)		0.00
<i>ddi</i>	**	-0.09 
<i>pul</i>	*	0.16 
<i>cnr</i>	***	0.22 
<i>bc</i>	.	0.06 
Adjusted R^2		0.04
p-value		0.00
Moran's I		0.04
p-value		0.05

the spatial information in the model. The adjusted R^2 rises to 0.47 and the spatial autocorrelation is very close to zero (i.e., Moran's $I = 0.06$, p -value = 0.01). At this point, I check what OA values of Foursquare or OSM categories of amenities are related to the residuals of the SAR model, through the method by Clifford *et al.* and False Discovery Rate (FDR). I find that no OA values is associated with such residuals. I thus perform a stepwise regression with the metrics of the configuration of urban form, as independent variables, and the residuals of the SAR model, as dependent one. Outcomes show that there is a significant, although weak, relationship (adjusted $R^2 = 0.04$) between the metrics selected by the stepwise procedure and life expectancy, controlled for income deprivation (see Table 6.4). In case of absolute independence of the residuals from the predictors and of linear dependency between the two, an explanatory power of 4% for the model on residuals would correspond to an overall explanatory power of 7.5%, if the former value is considered in relation to the proportion of information that is not explained by the first model (i.e., 53%). However, these conditions would be very difficult to meet as inter-dependencies are likely to be present. As a result, the fit of this model is quite modest. For what concerns the four coefficients selected by the stepwise technique, these are all statistically significant, at 90% confidence level, and are: Density of

Dead-end Intersections (*ddi*, $\beta = -0.09$), Percentage of Unbuilt Land (*pul*, $\beta = 0.16$), Connected Node Ratio (*cnr*, $\beta = 0.22$), and Betweenness Centrality (*bc*, $\beta = 0.06$). Finally, I perform the Moran's test on the residuals of this regression. Outputs of such test show a very weak statistical evidence of the presence of spatial autocorrelation (i.e., Moran's $I = 0.04$, p -value = 0.05). The model can thus be considered reliable.

Table 6.5: LR and SAR models for childhood obesity. Red bar means positively associated with childhood obesity, blue bar means negatively associated with childhood obesity. '.' significant at $p < 0.1$; '*' significant at $p < 0.05$; '**' significant at $p < 0.01$; '***' significant at $p < 0.001$. Income deprivation can explain more than half (i.e., 58%) of the variation of childhood obesity.







<i>Independent variable</i>	<i>LR</i>		<i>SAR</i>	
	<i>p</i> -value	β	<i>p</i> -value	β
(intercept)		0.00		-0.01
<i>income deprivation</i>	***	-0.69 	***	-0.49 
Adjusted R^2		0.48		0.58
<i>p</i> -value		0.00		0.00
Moran's I		0.33		0.04
<i>p</i> -value		0.00		0.05

Table 6.6: LR model for the residuals of the regression between income deprivation and childhood obesity. Red bar means positively associated with childhood obesity, blue bar means negatively associated with childhood obesity. '.' significant at $p < 0.1$; '*' significant at $p < 0.05$; '**' significant at $p < 0.01$; '***' significant at $p < 0.001$. Urban form can only explain up to 3% of the variance of the residuals of the regression between income deprivation and childhood obesity. The strongest coefficients are *id* (i.e., $\beta = -0.14$) and *oahp* (i.e., $\beta = 0.09$).

<i>Independent variable</i>	<i>LR</i>	
	<i>p</i> -value	β
(intercept)		0.00
<i>pga</i>	***	0.10 
<i>id</i>		-0.14 
<i>pul</i>	**	-0.08 
<i>oahp</i>	**	0.09 
Adjusted R^2		0.03
<i>p</i> -value		0.00
Moran's I		0.01
<i>p</i> -value		0.27

Outcomes of the linear regression with income deprivation as independent variable and childhood obesity as dependent one show that the former can explain 48% (adjusted $R^2 = 0.48$) of the variance of the latter, with an high statistical signifi-

cance (i.e., 99%) (see Table 6.5). I perform the Moran's test to check for spatial autocorrelation in the residuals and find statistical evidence of the presence of such phenomenon (i.e., Moran's $I = 0.33$, p -value = 0.00). I thus use the SAR technique to incorporate the spatial pattern of the observations into the equation. The adjusted R^2 rises to 0.58, meaning that income deprivation and the spatial factor can explain 58% of the variance of childhood obesity. The regression coefficient for income deprivation slightly decreases (i.e., from $\beta = -0.69$ to $\beta = -0.49$) due to the effect of including the spatial information in the model. The Moran's test performed on the residuals shows only weak statistical evidence of the presence of spatial autocorrelation (i.e., Moran's $I = 0.04$, p -value = 0.05). At this point, I use the method by Clifford *et al.* and the False Discovery Rate (FDR) technique to identify what OA values of Foursquare and OSM categories of amenities bear significant relationships with the residuals of the SAR model. Outputs suggest that none of the OA values are associated with the residuals. I thus perform a stepwise linear regression with the metrics of the configuration of urban form as independent variables and the residuals of the SAR model as dependent one. The outcomes of this regression show that the variables selected by the stepwise technique, Percentage of Green Areas (pga), Intersection Density (id), Percentage of Unbuilt Land (pul), and Offering Advantage of Historic Properties ($oahp$), can explain 3% of the variance of the residuals (see Table 6.6). Again, in an optimal scenario of independence of residuals from predictors and perfect linear relationship between the two, this would correspond to an overall explanatory power of 7%, when considering the proportion of information on childhood obesity that income deprivation does not explain (i.e., 42%). However, these ideal conditions would probably be hard to meet due to possible cross-dependencies between aspects of urban form and income, so results are rather modest. Three regression coefficients out of four are statistically significant at 99% confidence level: Percentage of Green Areas ($\beta = 0.10$), Percentage of Unbuilt Land ($\beta = -0.08$), and Offering Advantage of Historic Properties ($\beta = 0.09$). Intersection Density is not statistically significant, instead. The Moran's test on the residuals of this last regression does not show statistical evidence of the presence

of spatial autocorrelation (i.e., Moran's $I = 0.01$, p -value = 0.27). Results are thus reliable.

6.4.4 Interpretations

For what concerns life expectancy, results suggest that the typical London neighbourhood that promotes longer life spans is characterised by less Density of Dead-end Intersections (*ddi*), more Percentage of Unbuilt Land (*pul*), more Connected Node Ratio (*cnr*), and more Betweenness Centrality (*bc*). This means fewer cul-de-sac, more unbuilt areas, higher street network connectivity, and more accessibility to the wider city-scale network. Such features seem to depict the historic suburban neighbourhoods of London (e.g., Hampstead, Peckham): walk-able but not densely built urban environments with easy access to the wider city network. Such features might be linked to more life expectancy as a connected and walk-able urban environments may invite residents to have healthier life styles, for example, one that includes more walking than driving. Findings of this quantitative analysis seem to be in line with studies that examined the link between walk-able, connected neighbourhoods and health. Riggs and Gilderbloom, for example, found a significant relationship between years of potential life lost, computed as the difference between life expectancy and the age at which an individual actually die, and the connectivity and walk-ability of neighbourhoods of Louisville, KY [108]. Such effects were even stronger in areas hosting ethnic minorities and poor residents, confirming also the link between income and years of potential life lost and thus part of the findings of my analysis. Other researchers investigated the relationships between a walk-ability index, computed by considering street network connectivity among other factors, and a set of health measures in King County, WA. Findings suggested that a 5% increase in walk-ability corresponded to a per capita 32.1% increase in time dedicated to physically active travel and to a 0.23 point reduction in body mass index [113]. Watts *et al.* found that walk-able, connected neighbourhoods also had positive effects on cognitive health in the elderly [114].

As for childhood obesity, outcomes suggest that the typical London neighbourhood with lower levels of such phenomenon is characterised by more Percentage of

Green Areas (*pga*), less Percentage of Unbuilt Land (*pul*), and more Offering Advantage of Historic Properties (*oahp*). This corresponds to an urban environment characterized by parks and gardens plugged into the historic and relatively dense city fabric. These features might be linked to less childhood obesity as green and dense neighbourhoods might promote more physical activity, for example, in the forms of walking, running and outdoor playing. Several studies seem to back up my findings. Liu *et al.*, for example, performed a cross-sectional study of 7,334 subjects, aged between 3 and 18 years, in Marion County, IN, and found that, after controlling for socio-economic status, increased neighbourhood vegetation and higher population density were significantly associated with less overweight youth [109]. In a more recent study, performed in the same geographic context, Bell *et al.* reported that both greenery and density were associated with lower values of body mass index in young kids, however, the predictor for residential density performed more poorly [115]. The majority of published works, though, purely focused on the relationship between amount of, or proximity to, greenery and childhood obesity. In a recent literature review, Lachowycz and Jones found that 41 studies out of 60 reported the existence of an inverse relationship between these two phenomena [116].

6.5 Discussion

In this chapter, I used the general methodology presented in this thesis to analyse the relationship between amenities, configurational aspects of urban form and levels of deprivation, life expectancy, and childhood obesity in Greater London. Metrics to quantify urban form and the presence of amenities in neighbourhoods were extracted from openly accessible datasets, such as OS VectorMap District and OSM. Deprivation levels were obtained from the 2011 IMD dataset, while life expectancy and childhood obesity were extracted from the London Ward Well-Being Scores dataset. Through a regression analysis that accounted for spatial dependencies, I modelled the relationships between amenities, aspects of the configuration of urban form and each of the indexes considered in this analysis. Outcomes suggest that a

combination of specific categories of amenities and configurational features of urban form can explain 76% of the variance of deprivation levels. All the regression coefficients, except one (i.e., Betweenness Centrality), showed statistical significance, with Population Density, Density of Dead-end Intersections, Intersection Density, and Offering Advantage of Dentist's Offices showing the greatest coefficients. The first two showed positive signs, while the second two negative ones. The typical London neighbourhood with low levels of deprivation seem thus to have the features of the traditional compact city form: connectivity, few cul-de-sac, and historic urban fabric. Moreover, such neighbourhoods seem also to be characterized by the absence of amenities that offer unhealthy food and by the presence of sport facilities and amenities that offer surplus goods. As for the model for life expectancy, none of the metrics that measure the presence of amenities passed the filtering required by the analysis, thus only metrics of the configuration of urban form were used in the regression. After controlling for income deprivation, the model showed a weak explanatory power (at most 4%), with the strongest coefficients being Connected Node Ratio and Percentage of Unbuilt Land, both positive. Findings suggest that London neighbourhoods with more life expectancy are characterised by street network connectivity, few cul-de-sac, accessibility to the wider city network, and less built area. For what concerns the model for childhood obesity, none of the metrics that quantify the offering of amenities passed the filtering required by this analysis, thus only measures of the configuration of urban form were used in the regression. Having controlled for income deprivation, the model showed a weak adjusted R^2 of 0.03. The regression coefficients with the greatest values were Percentage of Green Areas and Offering Advantage of Historic Properties, both showing positive signs. Outcomes suggest that London neighbourhoods with low levels of childhood obesity are characterised by green areas plugged into the historic dense urban fabric.

6.5.1 Limitations

I ought to acknowledge several limitations for this investigation. First, while the model for IMD could explain a significant amount of variance of IMD (i.e., ad-

justed $R^2 = 0.76$), the ones for life expectancy and childhood obesity performed poorly, with adjusted R^2 values of 0.04 and 0.03, respectively. The interpretations derived from values and signs of their regression coefficients should be thus taken cautiously. One possible way to improve these outcomes would involve the use of other metrics of urban form (e.g., materials of buildings, distance to closest green area). Second, findings only hold for Greater London, for 2013, and are thus not generalisable. Finally, the outcomes of this analysis do not imply causation as the approach used is based on correlation and regression analyses.

In the next section, I sum up the content of this thesis, provide a full discussion of limitations, and present possible future work for developing the proposed method further.

Chapter 7

General Conclusions

In this last chapter, I summarise the contributions of this thesis and how these contributions can benefit different stakeholders. I then critically assess the work, by identifying limitations, and propose directions for future work.

7.1 Summary of Contributions

In this thesis, I proposed a quantitative methodology – potentially applicable to any geographic context – to test two hypotheses: whether amenities are related to aspects of city liveability and whether the configuration of the urban environment is associated with aspects of city liveability. Metrics to quantify such aspects are extracted from openly accessible datasets such as OpenStreetMap (OSM), Foursquare, and census data. The approach mainly consists in: (i) identifying theories focused on urban form and liveability; (ii) defining metrics, inspired by such theories, that function as proxies of urban form; (iii) computing such metrics; (iv) identifying metrics significantly associated with a liveability index, through correlation analysis; (v) performing regression analysis with the relevant metrics as independent variables and a liveability index as dependent one; (vi) interpreting quantitative findings in light of previous theories of urban form and outcomes of related disciplines. I present next a detailed breakdown of, and reflections about, the contributions I made.

7.1.1 Amenities and Socio-economic Deprivation

In Chapter 4, I applied the proposed methodology to test the first hypothesis of this thesis (i.e., whether amenities are related to aspects of city liveability). I modelled the relationship between the offering of categories of amenities in neighbourhoods and the deprivation levels of three UK cities, namely Greater London, Greater Manchester, and West Midlands. Information on amenities was extracted from Foursquare and OSM, while deprivation levels were extracted from the 2011 English Index of Multiple Deprivation (IMD). Outcomes suggested that some specific categories of amenities were related to IMD, across all case studies, and that these amenities could explain between 61% and 75% of the variance of such index, when accounting for the spatial autocorrelation associated with the variables. Through thematic analysis, I then derived five topics by looking at significant categories of amenities that were common across the three cities. These were: *surplus goods*, *sport & recreation*, *unhealthy food*, *unpleasant places*, and *public buildings*. The first two were associated with more advantaged neighbourhoods, while the last three were related to more deprived ones. These findings seem to corroborate several previous studies. In particular, the ones that found links between sport facilities, open space and wealth [16, 56, 93, 94, 95], the ones that reported associations between fast food restaurants and deprivation [17, 96, 97], and the ones that found a relationship between the presence of factories in neighbourhoods and deprivation [98].

7.1.2 Configuration of Urban Form and Socio-economic Deprivation

In Chapter 5, I used the proposed methodology presented in Chapter 3 to test the second hypothesis of this thesis (i.e., whether the configuration of the urban environment is associated with aspects of city liveability). I analysed the relationship between configurational aspects of urban form and deprivation levels of six UK cities: Greater London, Greater Manchester, West Midlands, Leeds, Liverpool, and Newcastle. Information on the configuration of urban form was extracted from OSM and OS VectorMap District, while the one on deprivation was obtained from

the 2011 IMD. Outcomes suggested the presence of a significant link between the configuration of urban form and deprivation. To be more specific, the models were able to explain between 27% and 70% of the variance of IMD. Values and signs of regression coefficients highlighted some patterns across the six cities. More deprived English neighbourhoods seemed to be characterised by a higher population density, larger unbuilt areas, more cul-de-sac, and a more regular street pattern. Such features resembled the modernist “towers in the park” design scheme (i.e., residential towers detached from side-walks, surrounded by open space). These findings seemed thus to corroborate theories supporting the compact city form (see, for example, [3, 32, 4]).

7.1.3 Amenities, Configuration of Urban Form and Three Liveability Indexes

In Chapter 6, I applied the methodology presented in Chapter 3 to test both hypotheses of this thesis (i.e., whether amenities are related to aspects of city liveability and whether the configuration of the urban environment is associated with aspects of city liveability) together. I studied the relationship between amenities, configuration of urban form and deprivation, life expectancy, and childhood obesity, in Greater London. Information on amenities was extracted from OSM and Foursquare, the one on the configuration of urban form from OSM and VectorMap District, while levels of the three indexes of liveability from the 2011 IMD, the 2013 life expectancy scores, and the 2013 childhood obesity scores, respectively. I computed three different models, one for each of the indexes considered. The model for IMD could explain 76% of the variance of the index. Outcomes of such model suggest that amenities offering *unhealthy food* and facilities associated with *unpleasant places* were associated with deprivation. Conversely, amenities offering *surplus goods* and *sport & recreation* were related to less deprived areas. Furthermore, aspects of the traditional compact city form (e.g., Intersection Density and Offering Advantage of Historic Buildings) were associated with more advantaged neighbourhoods, while more cul-de-sac and green areas were related to more deprived ones. The model for life expectancy could explain at most 4% of the index, after having controlled

for income. London neighbourhoods with a longer life expectancy seemed to be characterized by more connectivity, fewer cul-de-sac, more unbuilt areas, and more accessibility to the wider city network. Such features resembled historic suburban developments of London (e.g., Hampstead, Peckham). This finding seemed to corroborate previous studies (see, for example, [108, 113, 114]). The model for childhood obesity, controlled for income, could overall explain at most 3% of the variance of the index. London areas with less childhood obesity seemed to be characterised by the presence of more green areas plugged into a relatively dense historic urban fabric. I argued that such green areas might have been gardens or small parks, surrounded by buildings. This result seemed to be in line with previous research (see, for example, [109, 115, 116]).

7.2 Discussion

In this section, I first summarise the engineering work required to develop the content of this thesis. Second, I illustrate how one could put this work to practice. In particular, I show how to use the proposed metrics of urban form to predict deprivation, rather than only assessing their relationship.

7.2.1 Engineering

In developing this thesis, I acquired a varied set of skills. I deepened my knowledge of GIS, in particular, in importing and aggregating data, joining datasets, and producing maps. I learnt techniques of data analysis such as summary statistics, multiple testing control, correlation and regression analyses. I improved my coding skills in SQL, R, Java, and Python and wrote scripts to perform data extraction and computing tasks. These skills were fundamental to deal with the underlying engineering behind the outcomes of this thesis. This mainly consisted of the following processes:

- Data collection. While it was relatively easy to access and import information for most datasets, obtaining Foursquare data was not as straightforward. I thus had to adapt a Python script to query the Foursquare API and obtain the data.

- Pre-processing. This was necessary to select only the data associated with the study areas. Ordnance Survey data, for example, comes in tiles of 100 km by 100 km, which do not match the administrative boundaries of the cities under study. I thus had to use GIS to remove the unnecessary data. I had to follow a similar procedure with Foursquare data, as it was acquired through bounding boxes and thus did not match exactly the boundaries of the study areas.
- Data manipulation. This required various sub-steps and the use of different software. I used a script in Java, written by me and my colleague Giovanni Quattrone, to count the hundreds of Foursquare and OSM categories of amenities present in each spatial unit of analysis. I used a series of GIS commands to extract the node degrees from the street networks of the cities under study. This was necessary to then compute several of the metrics of the configuration of the built environment. Through GIS, I also computed the areas of building footprints and green areas for each unit of analysis. Finally, I used the Multiple Centrality Assessment (MCA) software to calculate the Betweenness Centrality values of the street networks of the cities under study.
- Computation of metrics. This task consisted in the implementation of the formulae for the computation of the proposed metrics. To do so, I used a script in Java, written by me and my colleague Giovanni Quattrone, to quickly compute the Offering Advantage values of hundreds of Foursquare and OSM categories. I used GIS to calculate the metrics of the configuration of the urban environment and produce choropleth maps.
- Data analysis. This was carried out mainly through R and various packages that allowed the analysis of frequency distributions, multiple testing control, multicollinearity check, and correlation and regression analyses in spatial context.

At the moment, the engineering presented above exists only as a collection of scripts and commands. This renders the analysis presented in this thesis hardly

accessible to anyone except the author of this thesis. It would be valuable to spend some engineering efforts to put all these scripts and commands in a web-based tool that could be used by other researchers too.

7.2.2 From Model to Prediction

Traditional methods for collecting census data (e.g., household surveys) are expensive and time consuming. As a result, such data is collected infrequently (from few years in developed countries to several in developing ones). As I illustrated in the previous chapters, the application of the proposed methodology to the UK cities confirmed the existence of significant relationships between aspects of urban form and IMD. To avoid performing household surveys and, instead, automatically estimate IMD from open data, one can try to build a classifier of such index by using the metrics of urban form that bear the strongest relationships. An experiment that I performed on Greater London indeed showed that this was possible. To carry out such experiment, I first randomly split the 625 London wards in train (i.e., 25%) and test (i.e., 75%) sets. Since a previous work, on the same city, predicted IMD for two bins (below and above the median value) [60], I subdivided the values of such index in the same exact bins. I then built a Naive Bayes classifier to estimate levels of deprivation of London wards that took in input the OA values of Foursquare and OSM categories, which were found to be significantly correlated with IMD in the analysis presented in Chapter 4. I present in Table 7.1 the performance of the classifier.

<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>	<i>Class</i>
0.763	0.692	0.726	50% more deprived
0.713	0.780	0.745	50% less deprived

Table 7.1: Performance of the classifier for Greater London, with IMD subdivided in two bins (below and above the median value).

To evaluate this model, I compared the results with the outcomes obtained by Smith *et al.* [60]. Although a precision of 0.763 for deprivation above median and of 0.713 for deprivation below median are lower than the best result reported by the cited work (i.e., 0.805), my model was still competitive as the outcomes obtained in the latter were valid only for 10% of London wards, while mine covered the whole

city. Furthermore, Smith *et al.* used telecommunication data to carry out their study, which is proprietary and thus difficult to retrieve for other areas or for different time frames. My work relied, instead, on the use of openly accessible data and can thus be potentially applied everywhere. For example, it can be used in Sub-Saharan African countries, where poverty indexes are very costly to compute.

7.3 Implications

Method and outcomes of this thesis potentially have both theoretical and practical implications. From a theoretical standpoint, researchers in the field of Urban Design and Social Science can use the proposed method to systematically analyse urban theories, for example, by first defining metrics inspired by such theories, and, second, by comparing these metrics to liveability indexes, across different countries and time frames. The same researchers can also use it to shed light on possible clashes among different urban theories. Moreover, they can utilise it to provide quantitative backgrounds to new theories, for example, by first defining metrics based on these new theories, and, second, by studying these metrics in relation to available liveability indexes.

For what concerns the outcomes of this thesis, they seem to validate previous works that favoured the traditional compact city form, for UK cities. Moreover, the link between connectivity and longer life expectancy and the one between green areas, in a relatively dense urban environment, and less childhood obesity seem to corroborate previous works carried out in a different geographic context (i.e., the US). See, for example, [108, 113, 114] for the former and [109, 115, 116] for the latter. Finally, findings for some specific Foursquare categories of amenities seem also to agree with, and add information to, previous studies. For example, the relationship found between more advantaged neighbourhoods and sport facilities seem to validate with previous works [16, 56].

From a practical perspective, the proposed approach can be used to build a tool for neighbourhood profiling based on the relationship between amenities, configuration of urban form and liveability levels. Such tool might be useful for various

stakeholders. Researchers might want to utilise it to systematically analyse urban theories, through a more user friendly interface. City administrators might want to use it to compare urban form and levels of liveability across different areas to understand what makes a neighbourhood liveable compared to what does not. Eventually, this information can inform the debate on current building policies or future ones. Neighbourhood associations might use the tool to discuss planned developments in light of the relationship between urban form and liveability associated with similar developments built in other areas. Finally, single citizens might use it to compare different parts of a city and select their favourite place to live.

7.4 Limitations

I ought to acknowledge several limitations for the methodology proposed in this thesis. First, it focuses on urban form only. Many other factors might affect aspects of city liveability, for example, specific housing policies, economic interventions, socio-cultural trends, migrations, gentrification processes. While identifying and including all the possible factors in a model would presumably be impossible, the knowledge from other fields can be useful to contextualise, criticise, and interpret the outcomes of the proposed methodology. Furthermore, the approach presented in this thesis mainly focuses on configurational aspects of urban form. However, these can potentially comprise of other elements such as building materials and properties of the facades.

A second limitation concerns the selection of the spatial unit of analysis. This choice inevitably comes with the issue of the Modifiable Areal Unit Problem (MAUP) [76], a source of potential bias associated with the selection of the spatial unit of analysis. In practice, metrics of spatial phenomena can vary substantially if computed for areas of different sizes or boundaries. While there is no systematic approach to solve this problem, one can try to test the robustness of his or her choice of spatial unit by (i) computing metrics for a specific unit of analysis, (ii) shifting the spatial units and recompute the metrics, (iii) checking whether the values of the metrics thus obtained are the same as (or similar to) the ones obtained at the first

step. Although, I did not carry out this procedure in this thesis, future works should perform this check too.

A third limitation regards the issue of spatial autocorrelation. While the majority of the models show ample statistical evidence of the absence of such phenomenon, a minority do show weak but significant statistical evidence of its presence. This might be associated with slightly inflated regression coefficients or instability of the signs. A possible solution for this would require testing alternative modelling techniques, such as the Getis Spatially Filtered Regression (GSFR) [117] (i.e., a regression technique that filters out the spatial component of each independent variable), or adding other relevant metrics, which would eventually reduce the remaining unexplained spatial variation.

A fourth limitation regards geographic and temporal generalizability. Outcomes of the proposed method are always bounded to specific geographic contexts and time frames and, for this reason, it is not possible to derive universal findings from them.

A fifth limitation concerns causality. The methodology presented has to be seen as an exploratory technique, which can be applied to test a variety of research questions concerning the urban environment and city liveability across different geographic contexts. The outputs of this analysis, in fact, do not establish the direction of causality between features of urban form and aspects of city liveability. This means, for example, that, although connectivity might be found to be related to well-being, this does not necessarily mean that increasing the connectivity of a place would correspond to an actual improvement of the liveability levels of that specific area. Grasping causality in this context would require, for example, some sort of A/B tests, which would involve making tangible interventions on parts A of the city, based on the outcomes of the model, and comparing how the effects of such changes relate to parts B of the city, where no interventions have been carried out.

A sixth limitation regards the assumption of linear relationships between the phenomena analysed. The proposed method, in fact, models relations in a linear fashion. However, it is possible that the relationship between aspects of urban form

and liveability are more complex. Interactions between some specific aspects of urban form might play important roles in explaining levels of liveability. For example, the combination of the OA values of fried chicken restaurants and pawn shops might better describe levels of deprivation in neighbourhoods than the two measures taken separately. Furthermore, there might be non-linear relationships among measures of urban form and liveability. For example, I illustrated in this chapter that higher intersection density corresponded to better socio-economic conditions. However, this might be true only until a certain threshold. Beyond this, intersection density might be detrimental. A possible future work involves testing whether variables of urban form have interactions among each other and whether non-linear relationships exist.

A final limitation concerns the multiple correlation testing of the OSM and Foursquare categories of amenities. Although the False Discovery Rate (FDR) technique ensures that the majority of false positives are discarded, statistical uncertainty remains. For example, it is possible that some of the categories of amenities considered associated with deprivation and used in the regressions were actually false positives. This might have introduced biases in the final outcomes. A possible solution to this would require the use of stricter methods, such as the Bonferroni test [118].

7.5 Future Work

Future work might span in several directions. Given the worldwide availability of OSM data, a possible line of work might investigate whether the proposed methodology can function only through the use of such source of geographic information. Ten years ago, this might not have been possible due the low geographic coverage of OSM. Today, though, this is no longer the case, especially in urbanised areas of developed countries. Furthermore, the proposed methodology can be applied not only to the UK cities presented in this thesis, but also to other geographic contexts. Given the unprecedented urbanisation trends affecting developing countries and their lack of information on poverty, the proposed method, with OSM data in input, can be

applied and be useful in these more challenging contexts too. Although OSM data coverage might be lower than in developed countries, the increasing diffusion of OSM and mapping activities for humanitarian purposes¹ might provide enough data to perform the analysis. If this were the case, one could, for example, analyse the relationship between features of the built environment and socio-economic levels for areas where this information exists and then use this knowledge to build a classifier of socio-economic deprivation for areas where this information is not available.

A further work might involve the creation of a toolkit for analysing neighbourhoods and levels of city liveability. This can be made through the integration of the various steps presented in Section 7.2.1 (Engineering) in one piece of software. Such tool might then be able to output detailed statistical analysis for the neighbourhoods of the selected city. This might include summary statistics for the metrics of the configuration of the urban environment, amenities, and liveability, frequency distributions, density distributions, and the outcomes of correlation and regression analyses. The tool might also provide graphic output, for example, choropleth maps to visualise how the metrics of urban form and amenities vary across a particular city.

To render the proposed methodology more robust, a possible line of work would consist in performing simulations on the data in input. For example, one might introduce random noise in the source data of a specific city, apply the proposed methodology to perform the analysis, and check whether the outcomes of the analysis with random noise are similar to the ones obtained from an analysis without any noise in the data.

The analyses performed for life expectancy and childhood obesity (see previous chapter) showed that the relationships between these and urban form are weak. This could be a starting point for further research. One can try to improve the outcomes of the models by implementing a multi-modal approach based on a public participation geographic information system (PPGIS) [119]. This technique consists in acquiring fine grained spatial knowledge through the dissemination of GIS

¹See, for example, <https://hotosm.org/>.

and mapping practices at the level of local groups. Such technique can thus provide quantitative in-depth local information, which paired with urban form, might improve the models for life expectancy and childhood obesity and provide further insights in health and urban studies.

Appendix A

A sample of the Index of Multiple Deprivation (IMD) dataset

Table A.1: Sample of the 2011 IMD dataset [27]. The LSOA CODE represents the identification code of each Lower-layer Super Output Area (LSOA). The LA CODE is the identification code of the larger administrative area. The IMD SCORE is the Index of Multiple Deprivation Score (the higher the score, the higher the deprivation of a LSOA). The RANK OF IMD SCORE is the Rank of the Index of Multiple Deprivation Score (the higher the rank score, the less deprived a LSOA).

<i>LSOA CODE</i>	<i>LA CODE</i>	<i>IMD SCORE</i>	<i>RANK OF IMD SCORE</i>
E01000001	00AA	6.161637	28814
E01000002	00AA	5.585054	29450
E01000003	00AA	13.29292	20047
E01000004	00AA	11.17135	22503
E01000005	00AA	21.36466	12978
E01000006	00AB	17.07765	16410
E01000007	00AB	37.23528	5438
E01000008	00AB	45.21677	3172
E01000009	00AB	29.41356	8462
E01000010	00AB	38.44798	5038
E01000011	00AB	29.48219	8421
E01000012	00AB	37.34236	5400
E01000013	00AB	39.79677	4626
E01000014	00AB	37.83172	5235
E01000015	00AB	38.81102	4926
E01000016	00AB	31.64162	7495
E01000017	00AB	35.34124	6099
E01000018	00AB	32.63291	7073
E01000019	00AB	32.84567	6993

Appendix B

Maps of the configuration of the urban environment and deprivation for the six urban areas under study

In the next pages, I present six figures, one for each of the urban areas considered in the analysis presented in Chapter 5. Each of these figures show ten maps, one for each of the metrics computed (i.e., nine of the configuration of the urban environment and one of deprivation). The figures come in the following order:

- **Figure B.1** represents the metrics for Newcastle;
- **Figure B.2** represents the metrics for Leeds;
- **Figure B.3** represents the metrics for Greater Manchester;
- **Figure B.4** represents the metrics for Liverpool;
- **Figure B.5** represents the metrics for West Midlands;
- **Figure B.6** represents the metrics for Greater London.

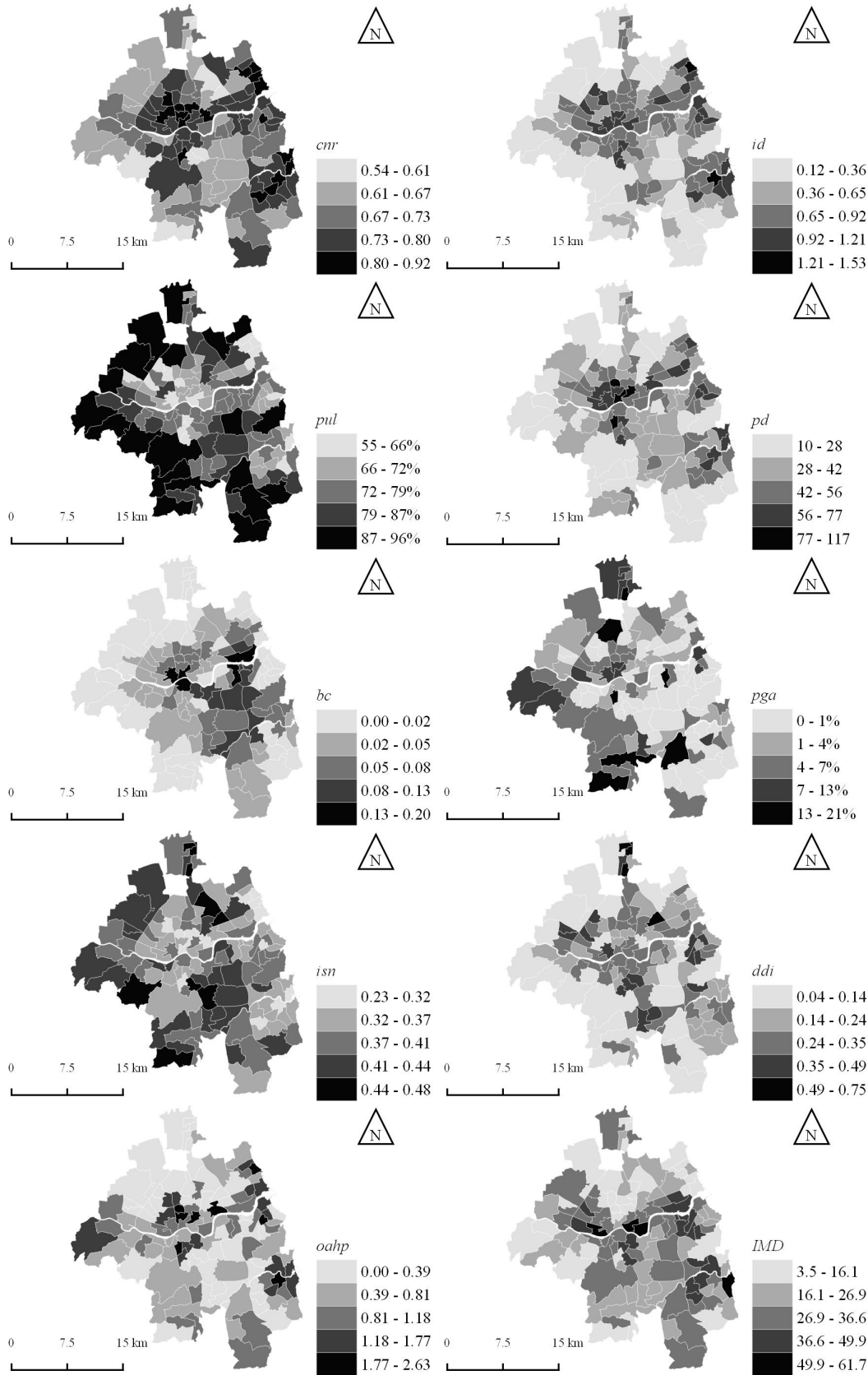


Figure B.1: Maps of the configuration of the urban environment and deprivation for Newcastle. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

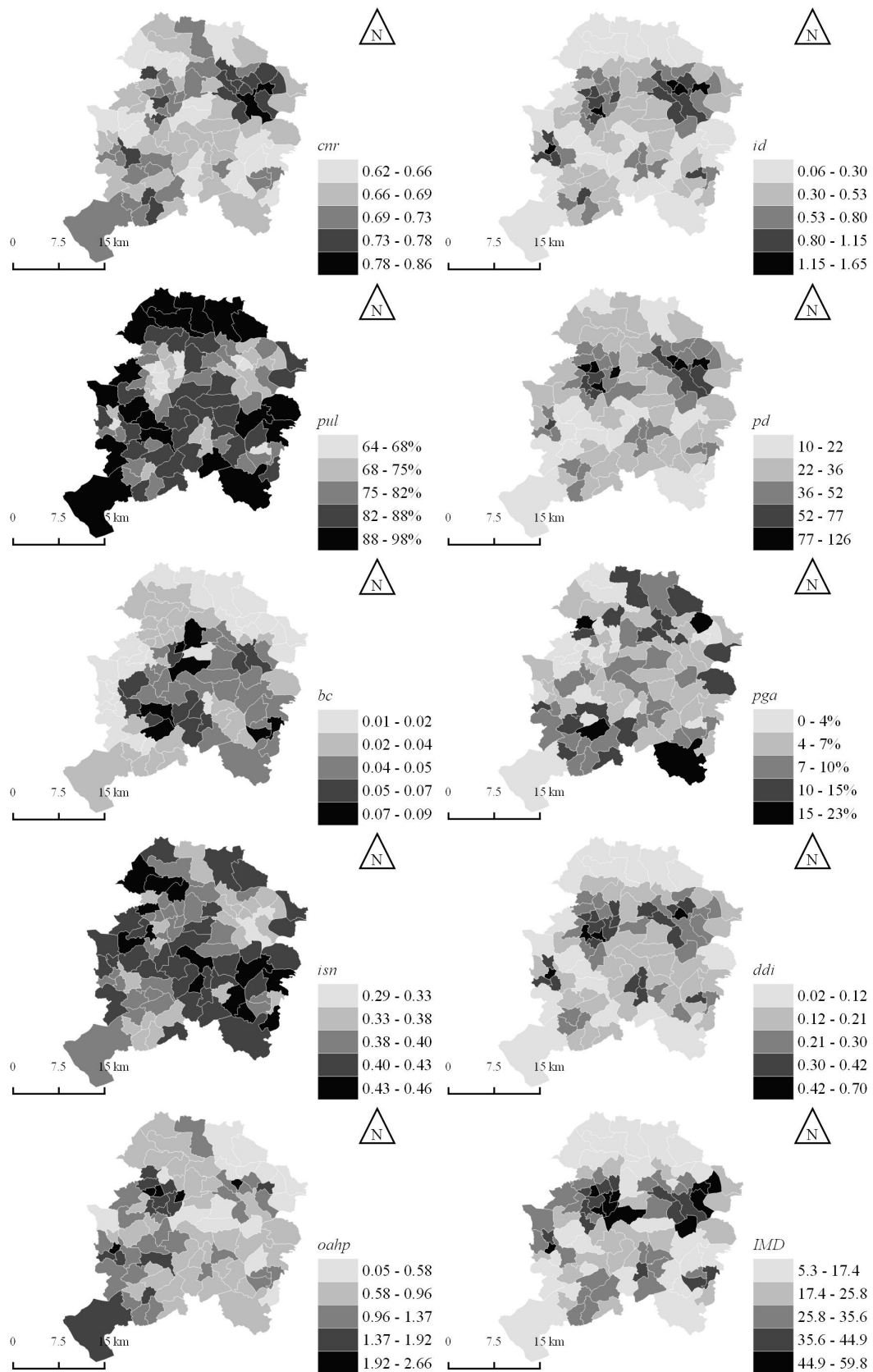


Figure B.2: Maps of the configuration of the urban environment and deprivation for Leeds. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

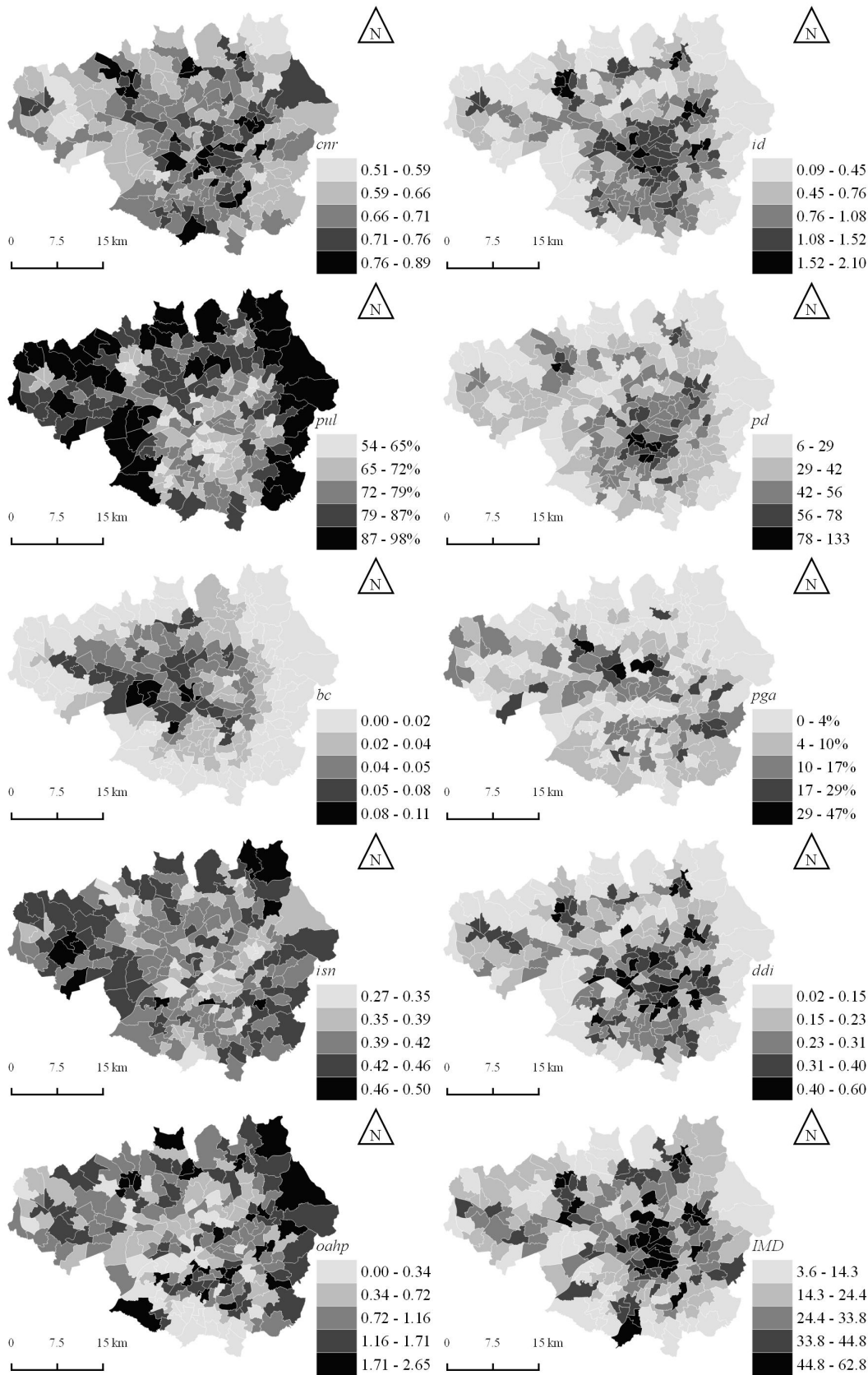


Figure B.3: Maps of the configuration of the urban environment and deprivation for Greater Manchester. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

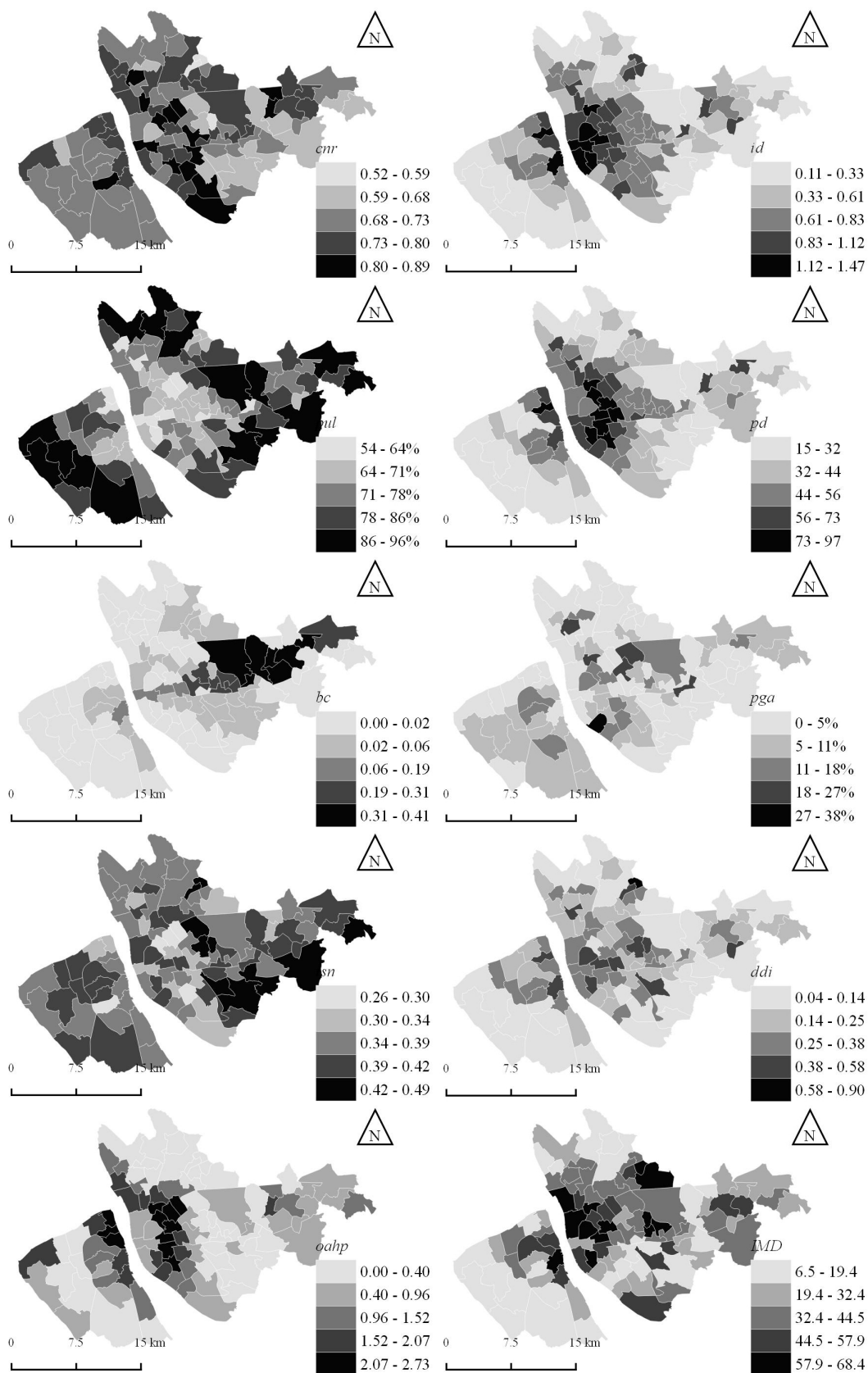


Figure B.4: Maps of the configuration of the urban environment and deprivation for Liverpool. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

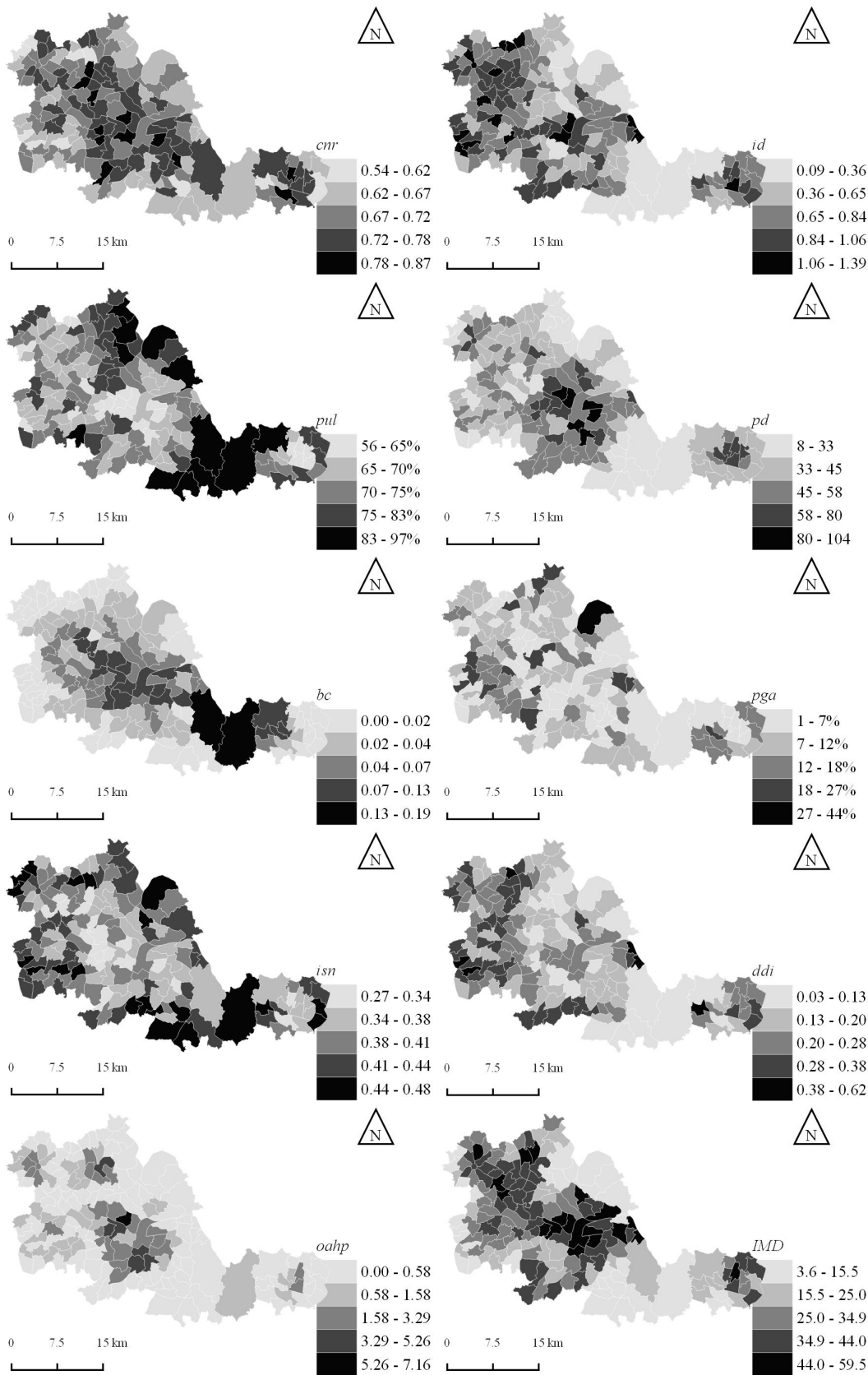


Figure B.5: Maps of the configuration of the urban environment and deprivation for West Midlands. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

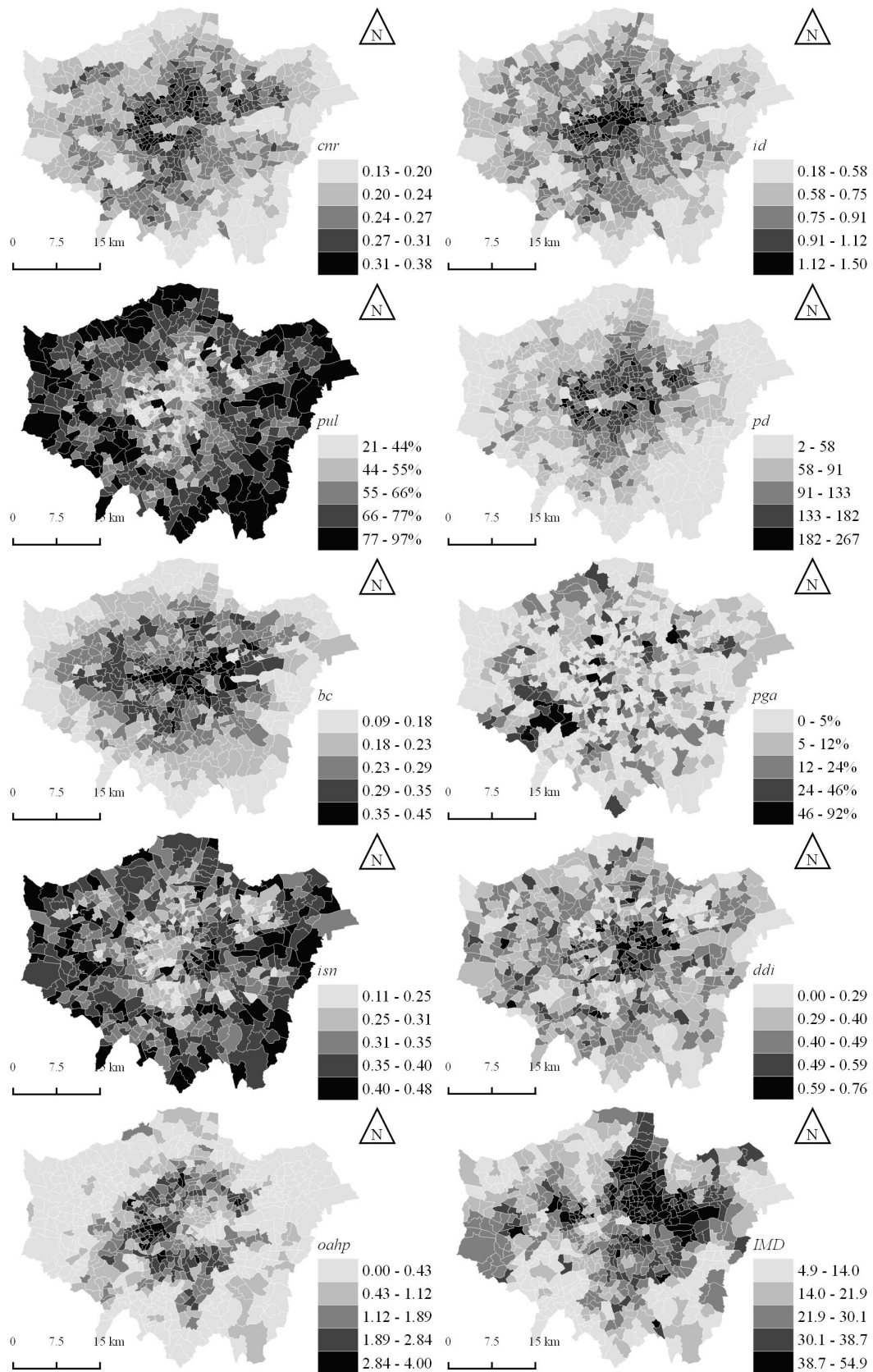


Figure B.6: Maps of the configuration of the urban environment and deprivation for Greater London. *cnr*: Connected Node Ratio. *id*: Intersection Density. *pul*: Percentage of Unbuilt Land. *bc*: Betweenness Centrality. *pga*: Percentage of Green Areas. *isn*: Irregularity of the Street Network. *ddi*: Density of Dead-end Intersections. *oahp*: Offering Advantage of Historic Properties. *IMD*: Index of Multiple Deprivation.

Bibliography

- [1] Saverio Muratori. *Studi per una operante storia urbana di Venezia*, volume 1. Istituto poligrafico dello Stato, Libreria dello Stato, 1959.
- [2] Michael Robert Gunter Conzen. Alnwick, Northumberland: a study in town-plan analysis. *Transactions and Papers (Institute of British Geographers)*, (27):iii–122, 1960.
- [3] Jane Jacobs. *The life and death of great American cities*. New York: Random House, 1961.
- [4] Jan Gehl. *Cities for people*. Island press, 2013.
- [5] Le Corbusier. *The city of tomorrow and its planning*. London: Architectural Press, 1947.
- [6] Ludwig Hilberseimer. *The new city: principles of planning*. Theobald, 1944.
- [7] DESA UN. World urbanization prospects, the 2011 revision: Highlights. *New York*. Online at http://esa.un.org/unpd/wup/pdf/WUP2011_Highlights.pdf, 2012.
- [8] Un-Habitat. *State of the World's Cities 2008-2009: Harmonious Cities*. Earthscan, 2008.
- [9] Bill Hillier. Can streets be made safe? *Urban design international*, 9(1):31–45, 2004.
- [10] Alan Geoffrey Wilson. *Complex spatial systems: the modelling foundations of urban and regional analysis*. Pearson Education, 2000.

- [11] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.
- [12] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences*, 104(17):7301–7306, 2007.
- [13] Emanuele Strano, Matheus Viana, Luciano da Fontoura Costa, Alessio Cardillo, Sergio Porta, and Vito Latora. Urban street networks, a comparative analysis of ten European cities. *Environment and Planning B: Planning and Design*, 40(6):1071–1086, 2013.
- [14] Viktor Mayer-Schönberger and Kenneth Cukier. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
- [15] Rob Kitchin. *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage, 2014.
- [16] Billie Giles-Corti and Robert J Donovan. Socioeconomic status differences in recreational physical activity levels and real and perceived access to a supportive physical environment. *Preventive medicine*, 35(6):601–611, 2002.
- [17] Steven CJ Cummins, Laura McKay, and Sally MacIntyre. McDonalds restaurants and neighborhood deprivation in Scotland and England. *American journal of preventive medicine*, 29(4):308–310, 2005.
- [18] Laura Vaughan, David L Chatford Clark, Ozlem Sahbaz, and Mordechai Muki Haklay. Space and exclusion: does urban morphology play a part in social deprivation? *Area*, 37(4):402–412, 2005.
- [19] Bill Hillier and Ozlem Sahbaz. An evidence based approach to crime and urban design, or, can we have vitality, sustainability and security all at once. *Bartlett School of Graduates Studies University College London*, 2008.

- [20] Martha Nussbaum and Amartya Sen. *The quality of life*. Oxford University Press, 1993.
- [21] Laza Kekic. The lottery of life. *The Economist*, Nov. 2012.
- [22] Mercer. 2015 quality of living survey. Mar. 2015.
- [23] United Nations Development Programme (UNDP). Human Development Report 2010. Nov. 2010.
- [24] Saamah Abdallah, Sam Thompson, Juliet Michaelson, Nic Marks, and Nicola Steuer. The Happy Planet Index 2.0: Why good lives don't have to cost the Earth. 2009.
- [25] Karma Ura, Sabina Alkire, Tshoki Zangmo, and Karma Wangdi. GNH Index. 2012.
- [26] Sabina Alkire, Adriana Conconi, and José Manuel Roche. Multidimensional Poverty Index 2012: Brief Methodological Note and Results. *University of oxford, Department of International Development, oxford Poverty and Human Development Initiative, oxford, UK*, 2012.
- [27] David McLennan, Helen Barnes, Michael Noble, Joanna Davies, Elisabeth Garratt, and Chris Dibben. The English indices of deprivation 2010. *London: Department for Communities and Local Government*, 2011.
- [28] Alfred Burns. Hippodamus and the planned city. *Historia: Zeitschrift für Alte Geschichte*, (H. 4):414–428, 1976.
- [29] Ferdinando Castagnoli. *Orthogonal town planning in antiquity*. The MIT Press, 1971.
- [30] Pollio Vitruvius, Valentinus Rose, and Hermann Müller-Strübing. *De architectura libri decem*. Teubneri, 1867.
- [31] Vincenzo Scamozzi. *L'idea della architettura universale*, volume 2. Colpo di fulmine Edizioni, 1907.

- [32] William H Whyte. *City: rediscovering its center*. New York: Doubleday, 1988.
- [33] Walter Gropius. *Scope of total architecture*. London: G. Allen & Unwin, 1956.
- [34] Ruth Glass. Introduction to London: Aspects of Change. London: Centre for Urban Studies, reprinted in R. Glass (ed.)(1989) *Clichés of Urban Doom*, 1963.
- [35] Michael Pacione. *Urban geography: A global perspective*. Routledge, 2009.
- [36] Tim Butler and Garry Robson. *London calling: the middle classes and the re-making of inner London*. Berg Publishers, 2003.
- [37] Jason Hackworth. Inner-city real estate investment, gentrification, and economic recession in New York City. *Environment and Planning A*, 33(5):863–880, 2001.
- [38] Mark Davidson and Loretta Lees. New-build gentrification and London's riverside renaissance. *Environment and planning A*, 37(7):1165–1190, 2005.
- [39] Kate Shaw et al. Culture, economics and evolution in gentrification. *Just Policy: A Journal of Australian Social Policy*, (28):42, 2002.
- [40] Bill Hillier. *Space is the Machine: A configurational theory of architecture*. 1996.
- [41] Oscar Newman. *Defensible space*. Macmillan New York, 1972.
- [42] Richard M Haughey. *Higher-density development: Myth and fact*. ULI–Urban Land Institute, 2005.
- [43] Keith Harries. Property crimes and violence in United States: An analysis of the influence of population density. *International Journal of Criminal Justice Sciences*, 1(2):24–34, 2006.

- [44] Jianling Li and Jack Rainwater. The real picture of land-use density and crime: a GIS application. In *Proceedings of the 20nd annual ESRI International User Conference*, 2000.
- [45] Andres Duany, Elizabeth Plater-Zyberk, and Jeff Speck. *Suburban nation: The rise of sprawl and the decline of the American dream*. Macmillan, 2001.
- [46] Emily Talen. Sense of community and neighbourhood form: An assessment of the social doctrine of new urbanism. *Urban studies*, 36(8):1361–1379, 1999.
- [47] Robert W Burchell, Naveed A Shad, David Listokin, Hilary Phillips, Anthony Downs, Samuel Seskin, Judy S Davis, Terry Moore, David Helton, and Michelle Gall. *The costs of sprawl-revisited*. Number Project H-10 FY’95. 1998.
- [48] Louis Wirth. Urbanism as a Way of Life. *American Journal of Sociology*, pages 1–24, 1938.
- [49] Georg Simmel. The metropolis and mental life. In Philip Kasnitz, editor, *Metropolis: Center and Symbol of our Times*, pages 30–45. Macmillan, Basingstonke, Hants, 2004.
- [50] Lance Freeman. The effects of sprawl on neighborhood social ties: An explanatory analysis. *Journal of the American Planning Association*, 67(1):69–77, 2001.
- [51] Ivonne Audirac and Maria Zifou. *Urban development issues: what is controversial in urban Sprawl?: an annotated bibliography of often-overlooked sources*. Number 247. Council of Planning Librarians, 1989.
- [52] AC Nelessen. *Visions for New American Dream: Process, Principles and an Ordinance to Plan and Design Small Communities*. Edward Brothers. 1994.

- [53] William B Shore. Recentralization: the single answer to more than a dozen United States problems and a major answer to poverty. *Journal of the American Planning Association*, 61(4):496–503, 1995.
- [54] Glen Bramley, Nicola Dempsey, Sinead Power, Caroline Brown, and David Watkins. Social sustainability and urban form: evidence from five British cities. *Environment and Planning A*, 41(9):2125–2142, 2009.
- [55] Glen Bramley and Sinead Power. Urban form and social sustainability: the role of density and housing type. *Environment and Planning B: Planning and Design*, 36(1):30–48, 2009.
- [56] Lisa M Powell, Sandy Slater, Frank J Chaloupka, and Deborah Harper. Availability of physical activity-related facilities and neighborhood demographic and socioeconomic characteristics: a national study. *American journal of public health*, 96(9):1676–1680, 2006.
- [57] Daniele Quercia, Neil Keith O’Hare, and Henriette Cramer. Aesthetic capital: what makes london look beautiful, quiet, and happy? In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*, pages 945–955. ACM, 2014.
- [58] Philip Saleses, Katja Schechtner, and César A Hidalgo. The collaborative image of the city: mapping the inequality of urban perception. *PloS one*, 8(7):e68400, 2013.
- [59] Nikhil Naik, Jade Philipoom, Ramesh Raskar, and César Hidalgo. Streetscore–Predicting the Perceived Safety of One Million Streetscapes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 793–799. IEEE, 2014.
- [60] C. Smith, D. Quercia, and L. Capra. Finger on the pulse: identifying deprivation using transit flow analysis. In *Proc. of CSCW*, pages 683–692. ACM, 2013.

- [61] Neal Lathia, Daniele Quercia, and Jon Crowcroft. The hidden image of the city: sensing community well-being from urban mobility. In *International Conference on Pervasive Computing*, pages 91–98. Springer, 2012.
- [62] Nathan Eagle, Michael Macy, and Rob Claxton. Network diversity and economic development. *Science*, 328(5981):1029–1031, 2010.
- [63] Huina Mao, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen. Mobile communications reveal the regional economy in Côte d’Ivoire. *Proc. of NetMob*, 2013.
- [64] Christopher Smith, Afra Mashhadi, and Licia Capra. Ubiquitous sensing for mapping poverty in developing countries. *Paper submitted to the Orange D4D Challenge*, 2013.
- [65] C. D. Elvidge, K. E. Baugh, E. A. Kihn, H. W. Kroehl, and E. R. Davis. Mapping city lights with nighttime data from the DMSP Operational Linescan System. *Photogrammetric engineering and Remote Sensing*, 63(6):727–734, 1997.
- [66] C. D Elvidge, M. L. Imhoff, K. E. Baugh, V. R. Hobson, I. Nelsonc, J. Safran, J. B. Dietz, and B. T. Tuttle. Night-time lights of the world: 1994-1995. *Photogrammetry and Remote Sensing*, 56(2):81–99, 2001.
- [67] A. M. Noor, V. A. Alegana, P. W. Gething, A. J. Tatem, and R. W. Snow. Using remotely sensed night-time light as a proxy for poverty in Africa. *Population Health Metrics*, 6(5):81–99, 2008.
- [68] Shlomo Angel. The Urban Expansion of London. <http://marroninstitute.nyu.edu/content/blog/the-urban-expansion-of-london/>, 2012. [Online; accessed 14-June-2017].
- [69] Bernard Porter. *Britannia’s Burden: the political evolution of modern Britain, 1851-1990*. E. Arnold, 1994.

- [70] Lionel Thomas Caswell Rolt. *Victorian engineering*, volume 1124. Penguin, 1974.
- [71] Harold James Dyos. The slums of victorian london. *Victorian Studies*, 11(1):5–40, 1967.
- [72] John Philip Ray. *The Night Blitz: 1940-1941*. Castle Books Incorporated, 2004.
- [73] John R Short. *Housing in Britain: the post-war experience*, volume 783. Taylor & Francis, 1982.
- [74] Nicholas Bullock. *Building the post-war world: modern architecture and reconstruction in Britain*. Psychology Press, 2002.
- [75] Greater London Authority. London Ward Well-Being Scores. <https://data.london.gov.uk/dataset/london-ward-well-being-scores>, 2014. [Online; accessed 25-March-2017].
- [76] Stan Openshaw and Peter J Taylor. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, 21:127–144, 1979.
- [77] Jerome L Myers, Arnold Well, and Robert Frederick Lorch. *Research design and statistical analysis*. Routledge, 2010.
- [78] Janne Lindqvist, Justin Cranshaw, Jason Wiese, Jason Hong, and John Zimmerman. I’m the mayor of my house: examining why people use foursquare-a social-driven location sharing application. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2409–2418. ACM, 2011.
- [79] Mordechai Haklay. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design*, 37(4):682–703, 2010.

- [80] Jean-François Girres and Guillaume Touya. Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 14(4):435–459, 2010.
- [81] Ina Ludwig, Angi Voss, and Maike Krause-Traudes. A Comparison of the Street Networks of Navteq and OSM in Germany. In *Advancing Geoinformation Science for a Changing World*, pages 65–84. Springer, 2011.
- [82] Błażej Ciepluch, Ricky Jacob, Peter Mooney, and Adam C Winstanley. Comparison of the accuracy of OpenStreetMap for Ireland with Google Maps and Bing Maps. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences 20-23rd July 2010*, page 337. University of Leicester, 2010.
- [83] César A Hidalgo, Bailey Klinger, A-L Barabási, and Ricardo Hausmann. The product space conditions the development of nations. *Science*, 317(5837):482–487, 2007.
- [84] R. Grenyer, C. D. L. Orme, S. F. Jackson, G. H. Thomas, R. G. Davies, T. J. Davies, K. E. Jones, V. A. Olson, R. S. Ridgely, P. C. Rasmussen, T. Ding, P. M. Bennett, T. M. Blackburn, K. J. Gaston, J. L. Gittleman, and I. P. F. Owens. Global distribution and conservation of rare and threatened vertebrates. *Nature*, 444:93–96, 2006.
- [85] Peter Clifford, Sylvia Richardson, and Denis Hémon. Assessing the significance of the correlation between two spatial processes. *Biometrics*, pages 123–134, 1989.
- [86] John D Storey and Robert Tibshirani. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100(16):9440–9445, 2003.
- [87] Craig M Bennett, George L Wolford, and Michael B Miller. The principled control of false positives in neuroimaging. *Social cognitive and affective neuroscience*, 4(4):417–422, 2009.

- [88] Samprit Chatterjee and Ali S Hadi. *Regression analysis by example*. John Wiley & Sons, 2015.
- [89] Andrew David Cliff and J Keith Ord. *Spatial processes: models & applications*. Taylor & Francis, 1981.
- [90] Robert P Haining. *Spatial data analysis: theory and practice*. Cambridge University Press, 2003.
- [91] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.
- [92] D Locker. Deprivation and oral health: a review. *Community dentistry and oral epidemiology*, 28(3):161–169, 2000.
- [93] Joke Luttk. The value of trees, water and open space as reflected by house prices in the Netherlands. *Landscape and urban planning*, 48(3):161–167, 2000.
- [94] Jacqueline Geoghegan. The value of open spaces in residential land use. *Land use policy*, 19(1):91–98, 2002.
- [95] Benjamin Bolitzer and Noelwah R Netusil. The impact of open spaces on property values in Portland, Oregon. *Journal of environmental management*, 59(3):185–193, 2000.
- [96] Jamie Pearce, Tony Blakely, Karen Witten, and Phil Bartie. Neighborhood deprivation and access to fast-food retailing: a national study. *American journal of preventive medicine*, 32(5):375–382, 2007.
- [97] Jason P Block, Richard A Scribner, and Karen B DeSalvo. Fast food, race/ethnicity, and income: a geographic analysis. *American journal of preventive medicine*, 27(3):211–217, 2004.
- [98] Susan A Perlin, David Wong, and Ken Sexton. Residential proximity to industrial sources of air pollution: interrelationships among race, poverty,

- and age. *Journal of the Air & Waste Management Association*, 51(3):406–421, 2001.
- [99] Afra Mashhadi, Giovanni Quattrone, and Licia Capra. Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 2013 conference on Computer supported cooperative work*, pages 611–622. ACM, 2013.
- [100] Giovanni Quattrone, Licia Capra, and Pasquale De Meo. There’s No Such Thing as the Perfect Map: Quantifying Bias in Spatial Crowd-sourcing Datasets. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1021–1032. ACM, 2015.
- [101] Norman W Garrick and Wesley E Marshall. The Shape of Sustainable Street Networks for Neighborhoods and Cities. In *Congress for New Urbanism*, pages 1–14, 2009.
- [102] Sergio Porta, Paolo Crucitti, and Vito Latora. The network analysis of urban streets: a primal approach. *Environment and Planning B: planning and design*, 33(5):705–725, 2006.
- [103] Fahui Wang, Anzhelika Antipova, and Sergio Porta. Street centrality and land use intensity in Baton Rouge, Louisiana. *Journal of Transport Geography*, 19(2):285–293, 2011.
- [104] Produit Timothée, Lachance-Bernard Nicolas, Strano Emanuele, Porta Sergio, and Joost Stéphane. A network based kernel density estimator applied to Barcelona economic activities. In *International Conference on Computational Science and Its Applications*, pages 32–45. Springer, 2010.
- [105] Adel Mohammad Remali, Sergio Porta, and Ombretta Romice. Correlating street quality, street life and street centrality in Tripoli, Libya. *The Past, Present and Future of High Streets*, 2014.

- [106] Yang Xiao. *Urban morphology and housing market*. PhD thesis, University College London, 2012.
- [107] Jolanda Maas, Robert A Verheij, Peter P Groenewegen, Sjerp De Vries, and Peter Spreeuwenberg. Green space, urbanity, and health: how strong is the relation? *Journal of epidemiology and community health*, 60(7):587–592, 2006.
- [108] William W Riggs and John Gilderbloom. The connection between neighborhood walkability and life longevity in a midsized city. *Focus*, 13(1):11, 2016.
- [109] Gilbert C Liu, Jeffrey S Wilson, Rong Qi, and Jun Ying. Green neighborhoods, food retail and childhood overweight: differences by population density. *American Journal of Health Promotion*, 21(4 suppl):317–325, 2007.
- [110] Gerry B Rodgers. Income and inequality as determinants of mortality: an international cross-section analysis. *Population studies*, 33(2):343–351, 1979.
- [111] Youfa Wang. Cross-national comparison of childhood obesity: the epidemic and the relationship between obesity and socioeconomic status. *International journal of epidemiology*, 30(5):1129–1136, 2001.
- [112] Allan J Brimicombe. Ethnicity, religion, and residential segregation in London: evidence from a computational typology of minority communities. *Environment and Planning B: Planning and Design*, 34(5):884–904, 2007.
- [113] Lawrence D Frank, James F Sallis, Terry L Conway, James E Chapman, Brian E Saelens, and William Bachman. Many pathways from land use to health: associations between neighborhood walkability and active transportation, body mass index, and air quality. *Journal of the American Planning Association*, 72(1):75–87, 2006.
- [114] Amber Watts, Farhana Ferdous, Keith Diaz Moore, and Jeffrey M Burns. Neighborhood integration and connectivity predict cognitive performance

and decline. *Gerontology and Geriatric Medicine*, 1:2333721415599141, 2015.

- [115] Janice F Bell, Jeffrey S Wilson, and Gilbert C Liu. Neighborhood greenness and 2-year changes in body mass index of children and youth. *American journal of preventive medicine*, 35(6):547–553, 2008.
- [116] K Lachowycz and AP Jones. Greenspace and obesity: a systematic review of the evidence. *Obesity reviews*, 12(5):e183–e189, 2011.
- [117] Luc Anselin and Arthur Getis. Spatial statistical analysis and geographic information systems. *The Annals of Regional Science*, 26(1):19–33, 1992.
- [118] Carlo E Bonferroni. *Teoria statistica delle classi e calcolo delle probabilita*. Libreria internazionale Seeber, 1936.
- [119] Nadine Schuurman. *GIS: a short introduction*, volume 4. Blackwell Publishing, 2004.